HTML

NIH Publication No: 02-4501

The Revised Up-and-Down Procedure:

A Test Method for Determining the Acute Oral Toxicity of Chemicals

Results of an Independent Peer Review Evaluation Organized by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)
and the
National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

Volume 1 of 2

National Institute of Environmental Health Sciences
National Institutes of Health
U.S. Public Health Service
Department of Health and Human Services

THE INTERAGENCY COORDINATING COMMITTEE ON THE VALIDATION OF ALTERNATIVE METHODS AND THE NTP INTERAGENCY CENTER FOR THE EVALUATION OF ALTERNATIVE TOXICOLOGICAL METHODS

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) was established in 1997 by the Director of the National Institute of Environmental Health Sciences (NIEHS) to implement NIEHS directives in Public Law 103-43. P.L. 103-43 directed NIEHS to develop and validate new test methods, and to establish criteria and processes for the validation and regulatory acceptance of toxicological testing methods. P. L. 106-545, the ICCVAM Authorization Act of 2000, established ICCVAM as a permanent committee. The Committee is composed of representatives from 15 Federal regulatory and research agencies and programs that generate, use, or provide information from toxicity test methods for risk assessment purposes. The Committee coordinates cross-agency issues relating to development, validation, acceptance, and national/international harmonization of toxicological test methods.

The National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods (Center) was established in 1998 to provide operational support for the ICCVAM, and to carry out committee-related activities such as peer reviews and workshops for test methods of interest to Federal agencies. The Center and ICCVAM coordinate the scientific review of the validation status of proposed methods and provide recommendations regarding their usefulness to appropriate agencies. The NTP Center and ICCVAM seek to promote the validation and regulatory acceptance of toxicological test methods that will enhance agencies' abilities to assess risks and make decisions, and that will refine, reduce, and replace animal use. The ultimate goal is the validation and regulatory acceptance of new test methods that are more predictive of human and ecological effects than currently available methods.

Additional Information

Additional information can be found at the ICCVAM/Center Website: http://iccvam.niehs.nih.gov and in the publication: *Validation and Regulatory Acceptance of Toxicological Test Methods, a Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods* (NIH Publication No. 97-3981, or you may contact the Center at telephone 919-541-3398, or by e-mail at iccvam@niehs.nih.gov. Specific questions about ICCVAM and the Center can be directed to the ICCVAM Co-chairs:

Dr. William S. Stokes, NIEHS, EC-17, P.O. Box 12233
Research Triangle Park, NC, 27709; 919-541-7997
stokes@niehs.nih.gov

Dr. Richard N. Hill, US EPA, MC-7101, 401 M Street, SW
Washington, DC, 20460; 202-260-2894
hill.richard@epa.gov

ICCVAM Agencies and Programs

Agency for Toxic Substances and Disease Registry
Consumer Product Safety Commission
Department of Agriculture
Department of Defense
Department of Energy
Department of Interior
Department of Transportation
Environmental Protection Agency
Food and Drug Administration
National Cancer Institute
National Institute of Environmental Health Sciences
National Institutes of Health, Office of the Director
National Institute of Occupational Safety and Health
National Library of Medicine
Occupational Safety and Health Administration

The Up-and-Down Procedure:
A Test Method For Determining the Acute Oral Toxicity of Chemicals

Results of an Independent Peer Review Evaluation Organized by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

Volume 1 of 2

National Toxicology Program
P.O. Box 12233
Research Triangle Park, NC 27709

November 2001
NIH Publication No. 02-4501

National Institute of Environmental Health Sciences
National Institutes of Health
U.S. Public Health Service
Department of Health and Human Services

Printed: 11/30/2001

Table of Contents

Volume 1 of 2

List of Abbreviations ..... **vii**

Peer Review Panel Members ..... **ix**

Acknowledgements ..... **x**

Preface ..... **xv**

Executive Summary ..... **xix**

I. Up-and-Down Procedure (UDP) Peer Panel Report, July 25, 2000 Meeting **I-1**

1. Introduction I-3
2. UDP Protocol: General Considerations I-6
3. UDP Protocol: Revised Primary Test I-10
4. UDP Protocol: Revised Limit Test I-14
5. UDP Protocol: Supplemental Test for Slope and Confidence Limits I-20
6. References I-25

Addendum I: Estimating a Point on the Dose-Response Curve I-29
Addendum II: Considerations for Estimating the Slope I-30
Addendum III: Summary of the Statistical Evaluation of the Revised UDP I-31
Addendum IV: Modified Isotonic Estimates of the Dose-Response Function I-33

II. UDP Peer Panel Report, August 21, 2001 Meeting **II-1**

1. Introduction II-3
2. Revised Draft UDP Test Guideline (July 12, 2001) II-4
3. Proposed Confidence Interval Procedure II-7
4. Software Program II-8
5. References II-9

III. References **III-1**

Appendices

A - ICCVAM Test Method Recommendations.....A-1

B - Final Revised UDP Test Guideline (U.S. EPA 10/31/01).....B-1

C - August 2001 Peer Panel Meeting Review Materials C-1
C-1 Acute Oral Toxicity: Revised UDP Test Guideline, July 12, 2001 .....C-3
C-2 A Proposed Procedure for Calculating Confidence Intervals C-33
C-3 Description of the Acute Oral Toxicity Software Program.....C-109
C-4 Evaluation Guidance to the Panel for the Revised UDP.....C-111

D - *Federal Register* Notices for the UDP D-1
D-1 Vol. 65, No. 34, February 18, 2000.....D-3
Request for Data and Nomination of Expert Scientists
D-2 Vol. 65, No. 106, June 1, 2000.....D-5
Notice of Peer Review Panel Meeting and Request for Comments
D-3 Vol. 66, No. 121, June 22, 2001.....D-7
Notice of Availability and Request for Comments
D-4 Vol. 66, No. 133, July 21, 2001 .....D-11
Notice of Peer Review Panel Meeting

E - Summary Minutes and Public Comments from the UDP Meetings.....E-1
E-1 Minutes and Public Comments of the Peer Review Panel Meeting..... E-3
July 25, 2000 in Crystal City, Arlington, VA
E-2 Minutes of the Peer Review Panel Meeting..... E-13
August 21, 2001 in Research Triangle Park, NC
E-3 Submitted Public Comment for the August 21, 2001 Meeting..... E-19

Table of Contents

Volume 2 of 2

Note: The April 14, 2000 Background Review Document (BRD), evaluated for the July 25, 2000 Peer Review Panel Meeting, is available at http://iccvam.niehs.nih.gov/methods/udpdocs/AllBRDlk.pdf. For this final report, the April 14, 2000 BRD was revised to address comments from the Panel (Appendix F). Selected BRD appendices from the April 14, 2000 BRD are provided for reference (Appendices G-Q).

F - Revised Background Review Document (BRD) on the UDP (October 31, 2001) F-1

G - Acute Oral Toxicity: Modified UDP G-1
(U.S. EPA Revised Test Guideline 425N, April 2000)

H - Acute Oral Toxicity: UDP H-1
(OECD Test Guideline 425, adopted 21 September 1998)

I - Acute Oral Toxicity I-1
(OECD Test Guideline 401, adopted 24 February 1987)

J - Development of OECD 425 .....J-1
J-1 UDP: Is there a Need for Further Validation?..... J-3
J-2 Rationale for the UDP as Submitted to OECD J-11
J-3 Presentation for Test Guideline 425 – UDP J-15

K - UDP Primary Test: Proposed Revision of the Guideline 425 .....K-1
"Primary Procedure" for Point Estimation of the LD50

L - UDP Primary Test: Comparison of Stopping Rules and LD50 Estimators..... L-1

M - The UDP Limit Test: Accuracy of *In Vivo* Limit Dose Tests M-1

N - Proposed UDP Supplemental Procedure to Estimate Slope and Confidence Interval N-1
N-1 Considerations for Supplemental Procedure N-3
N-2 Supplemental Procedure to Determine Slope and Confidence Interval N-5
N-3 Summary Tables.....N-9
N-4 Simulation Tables and Legends N-23
N-5 Additional Simulations: Supplemental Procedures to Determine Slope.....N-105

O - The Basis for Revising the UDP.....O-1
O-1 Statistical Basis for Estimating Acute Oral Toxicity O-3
O-2 Comparison of Classification Probabilities.....O-13
O-3 Brief Statistical Description of the Method.....O-17

P - Selection of the Appropriate Animal Gender for the UDP..... P-1
P-1 Gender Sensitivity of Xenobiotics P-3
P-2 Comparison of Male and Female Rat Oral and Dermal LD50 Values P-23
P-3 Toxicological Evaluation of Pesticide Hazard to Avian Wildlife.....P-45
P-4 Sex Dependent Metabolism of Xenobiotics.....P-69

Q - U.S. Federal Regulations for Acute Oral Toxicity.....Q-1
Q-1 CPSC Regulations, Administration and Enforcement Q-3
Q-2 OSHA Regulations, Hazard Communication Q-11
Q-3 U.S. EPA Regulations, Pesticide Registration Q-17
Q-4 U.S. EPA Regulations, Pesticide Labeling.....Q-25
Q-5 U.S. EPA Regulations, Data Requirements Q-33
Q-6 U.S. EPA Regulations, New Uses.....Q-57
Q-7 U.S. DOT Regulations, Shipping Requirements.....Q-69

List of Abbreviations and Acronyms

ASTM	American Society for Testing and Materials
ATCM	Acute Toxic Class Method
BRD	Background Review Document
°C	Degrees Centigrade
CASRN	Chemical Abstract Service Registry Number
CFR	Code of Federal Regulations
CI	Confidence Interval
CIIT	CIIT Centers for Health Research (formerly: Chemical Industry Institute of Toxicology)
CPSC	Consumer Product Safety Commission
ECETOC	European Centre for Ecotoxicology and Toxicology of Chemicals
ECVAM	European Centre for the Validation of Alternative Methods
EU	European Union
FDA	Food and Drug Administration
FDP	Fixed-Dose Procedure
FIFRA	Federal Insecticide, Fungicide and Rodenticide Act
FR	*Federal Register*
g	gram
GHS	Globally Harmonized System
GLP	Good Laboratory Practice
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
IUCLID	International Uniform ChemicaL Information Database
kg	kilogram
LD50	Median lethal dose
MEIC	Multicentre Evaluation of *In Vitro* Cytotoxicity
mg	milligrams
mL	milliliter
NICEATM	NTP Interagency Center for the Evaluation of Alternative Toxicological Methods
NIEHS	National Institute of Environmental Health Sciences
NTP	National Toxicology Program
OECD	Organisation of Economic Co-operation and Development
OPP	Office of Pesticide Programs/U.S. EPA
OPPT	Office of Pollution Prevention and Toxics/U.S. EPA
OPPTS	Office of Prevention, Pesticides, and Toxic Substances/U.S. EPA
PL	Public Law
SAS	Statistical Analysis System – (SAS Institute, Inc., Cary, NC, USA)
TG	Test Guideline
TG 401	Test Guideline 401 (Acute Oral Toxicity) [OECD]
TG 420	Test Guideline 420 (Acute Oral Toxicity - Fixed Dose Method) [OECD]

List of Abbreviations and Acronyms
(continued)

TG 423	Test Guideline 423 (Acute Oral Toxicity - Acute Toxic Class Method) [OECD]
TG 425	Test Guideline 425 (Acute Oral Toxicity - Up-and-Down Procedure) [OECD]
UDP	Up-and-Down Procedure
U.S. DOT	U.S. Department of Transportation
U.S. EPA	U.S. Environmental Protection Agency
ZEBET	Center for Documentation and Evaluation of Alternative Methods to Animal Experiments
3Rs	Refinement, Reduction, and Replacement (of animal use)

Peer Review Panel

The following individuals served as members of the Up-and-Down Procedure Peer Review Panel:

Curtis Klaassen, Ph.D. (Co-Chair)
University of Kansas Medical Center
Kansas City, KS

Diane Gerken, D.V.M., Ph.D. (Co-Chair)
Battelle Memorial Institute
Columbus, OH

PANEL SECTIONS

1. Revised UDP Protocol: General Considerations

Janice Kuhn, Ph.D., D.A.B.T. (Leader)
Stillmeadow, Inc.
Sugarland, TX

Kimberly Bonnette, M.S., L.A.T.G.
Springborn Laboratories, Inc.
Spencerville, OH

Gary Wnorowski, B.S.
Product Safety Labs
East Brunswick, NJ

2. Revised UDP Primary Test

Wallace Hayes, Ph.D., D.A.B.T., D.A.T.S. (Leader)
The Gillette Company
Boston, MA

Bas Blaauboer, Ph.D.
Utrecht University
Utrecht, The Netherlands

Robert Copeland, Ph.D.
Howard University
Washington, DC

John Reeve, M.S.
Ministry of Agriculture and Forestry
Food Assurance Authority
Wellington, New Zealand

Nigel Stallard, Ph.D.
University of Reading
East Gate Reading, UK

3. Revised UDP Limit Test

George Alexeeff, Ph.D., D.A.B.T. (Leader)
California Environmental Protection Agency
Sacramento, CA

Robert Condon, Ph.D.
Consulting Biostatician
Myersville, MD

A.A.J. van Iersel, Ph.D.
RIVM-Institute's Centre for Alternatives to Animal Testing
National Institute for Public Health and the Environment
Roosendaal, The Netherlands

4. UDP Supplemental Test for Slope/ Confidence Limits

Robert Scala, Ph.D., D.A.B.T., D.A.T.S. (Leader)
Toxicology Consultant
Tucson, AZ

Philip Botham, Ph.D.
Zeneca Ltd.
Cheshire, UK

Wyman Dorough, Ph.D.
Mississippi State University
Starkville, MS

Nancy Fluornoy, Ph.D.
American University
Washington, DC

Charles Hastings, Ph.D., D.A.B.T.
BASF Corporation
Research Triangle Park, NC

Acknowledgements

The following individuals developed the revised test method protocol and supporting documentation for the Up-and-Down Procedure. Those marked with an astrisk (*) were members of the UDP Technical Task Force:

***Dr. Greg Carr**
Proctor & Gamble Company

Mr. Wolfgang Diener
Federal Institute for Health Protection of Consumers and Veterinary Medicine (BgVV)

***Mr. David Farrar**
U.S. EPA

***Dr. Michael Green**
CPSC

***Dr. Kailash Gupta**
CPSC

Dr. Elwood F. Hill
Patuxent Environmental Science Center

***Dr. Timothy Barry**
U.S. EPA

Dr. Gregory Kedderis
CIIT

***Dr. Elizabeth Margosches**
U.S. EPA

***Mr. John Redden**
U.S. EPA

***Dr. Deborah McCall**
U.S. EPA

***Mr. William Meyer**
U.S. EPA

Dr. Cheryl Mugford
CIIT

Dr. Harold Podall
U.S. EPA

Dr. Carrie Rabe
Clement International Corporation

***Dr. Amy Rispin (Leader)**
U.S. EPA

Dr. Sharon Segel
Clement International Corporation

***Dr. Katherine Stitzel**
Proctor & Gamble Company

Mrs. Anne Whitehead
University of Reading

The following individuals are acknowledged for their contribution to the Revised Up-and-Down Procedure review process:

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) Acute Toxicity Working Group (ATWG)

Agency for Toxic Substances and Disease Registry (ATSDR)

Dr. John Wheeler

Consumer Product Safety Commission (CPSC)

Dr. Marilyn Wind, Director
Dr. Kailash Gupta
Dr. Susan Aitken

Department of Defense (DOD)

Dr. Harry Salem

Department of Transportation (DOT)

Dr. George Cushmac

Food and Drug Administration (FDA)

Dr. Nakissa Sadrieh
Dr. Antonia Mattia
Dr. Patrick G. Swann
Ms. Suzanne Fitzpatrick

National Cancer Institute (NCI)

Dr. Victor A. Fung

National Institute of Environmental Health Sciences (NIEHS)

Dr. William S. Stokes (Co-Chair)
Dr. Rajendra Chhabra

National Institute for Occupational Safety and Health (NIOSH)

Dr. Joe Antonini
Dr. Surender Ahir

U.S. Environmental Protection Agency (U.S. EPA)

Dr. Richard Hill (Co-Chair)
Dr. Angela Auletta
Dr. Byron Backus
Dr. Diane Beal
Mr. David Farrar
Dr. Roger Gardner
Dr. Bentley Gregg
Dr. Karen Hamernik
Dr. Masih Hashim
Ms. Marianne Lewis
Dr. Elizabeth Margosches
Dr. Jeanie McAndrew
Dr. Debbie McCall
Dr. Mark Perry
Dr. John Redden
Dr. Daniel Rieder
Dr. Amy Rispin
Dr. Roy Sjoblad

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)

Designated Agency Representatives

Agency for Toxic Substances and Disease Registry
*William Cibulas, Ph.D
◊Moiz Mumtaz, Ph.D.

Consumer Product Safety Commission
*Marilyn L. Wind, Ph.D.
Susan Aitken, Ph.D.
Kailash C. Gupta, Ph.D.
Patricia Bittner

Department of Agriculture
*Jodie Kulpa, D.V.M.
◊Elizabeth Goldentyer, D.V.M.

Department of Defense
*Robert E. Foster, Ph.D.
◊Patty Boll
Harry Salem, Ph.D.
John M. Frazier, Ph.D.

Department of Energy
*Marvin E. Frazier, Ph.D.
◊Marvin Stodolsky, Ph.D.

Department of the Interior
*Barnett A. Rattner, Ph.D.
◊Sarah Gerould, Ph.D.

Department of Transportation
*George Cushmac, Ph.D.
◊Steve Hwang, Ph.D.

Environmental Protection Agency
*Vanessa Vu, Ph.D.
Office of Pollution Prevention and Toxics
Richard N. Hill, M.D., Ph.D. (Co-Chair)
Angela Auletta, Ph.D
Office of Pesticides Programs
Karen Hamernik, Ph.D.
Office of Research and Development
Harold Zenick, Ph.D.
Suzanne McMaster, Ph.D.
OECD Test Guidelines Program
Maurice Zeeman, Ph.D.

Food and Drug Administration
*Leonard M. Schechtman, Ph.D.
◊Suzanne C. Fitzpatrick, Ph.D., D.A.B.T.
Center for Drug Evaluation and Research
Joseph J. DeGeorge, Ph.D.
Abby C. Jacobs, Ph.D.
Center for Devices and Radiological Health
Raju Kammula., D.V.M., Ph.D., D.A.B.T.
Melvin E. Stratmeyer, Ph.D.
Center for Biologics Evaluation and Research
Anne M. Pilaro, Ph.D.
Martin D. Green, Ph.D.
Center for Food Safety and Nutrition
David G. Hattan, Ph.D.
Center for Veterinary Medicine
Devaraya Jagannath, Ph.D.
Louis T. Mulligan, D.V.M.
National Center for Toxicological Research
William T. Allaben, Ph.D.
Martha M. Moore, Ph.D.
Office of Regulatory Affairs
Atin R. Datta, Ph.D.

National Cancer Institute
*David G. Longfellow, Ph.D.
◊Alan Poland, Ph.D.

National Institute of Environmental Health Sciences
*William S. Stokes, D.V.M., D.A.C.L.A.M. (Co-Chair)
◊John R. Bucher, Ph.D., D.A.B.T.
Rajendra S. Chhabra, Ph.D., D.A.B.T
Jerrold J. Heindel, Ph.D.

National Institute for Occupational Safety and Health
*Paul Nicolaysen, V.M.D.
◊Douglas Sharpnack, D.V.M., M.S., D.A.C.V.P.

National Institutes of Health
*Margaret D. Snyder, Ph.D.
◊Nelson Garnett, D.V.M., D.A.C.L.A.M.

National Library of Medicine
*Vera Hudson, M.S.
◊Jeanne Goshorn, M.S.

Occupational Safety and Health Administration
*Surender Ahir, Ph.D.

---

* Principal Agency Representative
◊ Alternate Principal Agency Representative

National Toxicology Program (NTP)
Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

Mr. Brad Blackard
ILS, Inc.

Ms. Sue Brenzel
ILS, Inc.

Ms. Bonnie Carson
ILS, Inc.

Dr. Finis Cavender
ILS, Inc.

Ms. Loretta Frye
NIEHS

Dr. Thomas Goldsworthy
ILS, Inc.

Ms. Christina Inhof
ILS, Inc.

Ms. Linda Litchfield
ILS, Inc.

Dr. Barry Margolin
UNC-Chapel Hill

Ms. Debbie McCarley
NIEHS

Dr. William S. Stokes, Director
NIEHS

Dr. Raymond Tice
ILS, Inc.

Preface

In the past the testing of chemicals for acute toxicity focused on determining the dose level which killed half the animals (the median lethal dose, or LD50). The "classical" LD50 used up to 100 animals to determine a median lethal dose within certain statistical bounds. More recently, several methods, which use far fewer animals, have been proposed and adopted. Attention has also broadened to include careful observation related to the onset, nature, severity, and reversibility of toxicity as well as lethality following single chemical exposures. Such information is crucial to properly identify, classify, and label human health hazards that may result from acute exposures in the workplace and home, and to make judgments pertaining to acute chemical hazards.

In 1981, the Organisation for Economic Co-operation and Development (OECD) adopted an international test guideline for acute oral toxicity testing which used as few as 30 animals. This guideline was revised in 1987 to reduce the number of test animals to as few as 20. In a continuing attempt to improve the estimate of acute toxicity while further reducing the number of animals used per test, three alternative test methods were subsequently developed and adopted as additional OECD Guidelines for acute toxicity. These were the Fixed Dose Procedure (FDP), the Acute Toxic Class Method (ATCM), and the Up-and-Down Procedure (UDP). Each of these methods used fewer animals when compared to the OECD 1987 conventional LD50 procedure.

In 1998, the OECD proposed deletion of the conventional LD50 test in light of the adoption of the three alternative methods (FDP, ATCM, UDP). Prior to formal deletion, the OECD determined it was necessary to revise the three methods to conform to a new globally harmonized hazard classification scheme. The U.S. EPA agreed to organize a Technical Task Force to revise the UDP. The Task Force was charged with preparing a revised UDP which comprised three procedures: a Primary Test to estimate the LD50, which would use an average of seven animals; a Limit Test for substances anticipated to have minimal toxicity; and a Supplemental Test to determine the slope and confidence interval for the dose-response curve. The Task Force used computer simulations to help revise the test. The revised UDP was proposed as a substitute for the existing conventional LD50 test (OECD Test Guideline 401, 1987; EPA OPPTS 870.1100, 1998) used to evaluate the acute oral toxicity potential of chemicals.

In August of 1999, the U.S. EPA asked the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) to conduct an independent scientific peer review evaluation of the revised UDP. Upon agreement, the ICCVAM requested knowledgeable individuals from participating Federal agencies to serve on an ICCVAM Acute Toxicity Working Group (ATWG); subsequently, ICCVAM would organize the peer review in collaboration with the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Methods (NICEATM). The ATWG held its first meeting in November 1999, and was charged with reviewing the revised UDP submission for completeness, proposing expert scientists for the peer review panel, developing questions for the peer review panel, and developing draft ICCVAM test recommendations based on the peer review panel's evaluation. During the next six months the ATWG provided guidance and interacted with the UDP Technical Task Force and ICCVAM to assemble adequate information for scientific peer review of the method in accordance with the *ICCVAM Test Method Submission Guidelines* (ICCVAM, 1999). A final revised UDP was submitted to ICCVAM in April 2000.

A *Federal Register* notice (February 18, 2000, Vol. 65, No. 34, 8385-8386) requested nominations of experts for the peer review panel (Panel). Nominations were also solicited from Federal agencies and national and international professional societies and organizations. The ATWG recommended a Panel composition with a broad range of experience and expertise, including

acute toxicity testing, biostatistics, alternative methods, pharmacology, and toxicokinetics. The Panel was composed of 19 experts from industry, academia, and government, and included scientists from the US, UK, New Zealand, and The Netherlands.

The Panel was charged with evaluating the usefulness and limitations of the three tests described in the UDP (Primary Test, Limit Test, and Supplemental Test) as a substitute for the conventional LD50. In reaching this determination, the Panel was asked to evaluate all available information and data on the UDP and to assess the extent to which each of the ICCVAM validation and regulatory acceptance criteria were addressed. These criteria are described in the document *Validation and Regulatory Acceptance of Toxicological Test Methods: A report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods*, NIH Publication No. 97-3981 (ICCVAM, 1997). A series of questions were posed to the Panel to facilitate ICCVAM and agency decisions on the UDP.

A request for data and information regarding the usefulness of the UDP, including information pertaining to completed, ongoing, or planned studies, was made via a *Federal Register* notice (February 18, 2000, Vol. 65, No. 34, 8385-8386). The availability of the UDP test method submission materials, a request for public comments, and announcement of the planned public peer review meeting were announced in a subsequent *Federal Register* notice (June 1, 2000, Vol. 65, No. 106, 35109-35110). All comments and information submitted in response to these notices were provided to the Panel in advance of the peer review meeting.

The Panel met in public session on July 25, 2000, in Arlington, Virginia. Panel members presented their evaluations and proposed conclusions and recommendations on each of the major sections and the Panel subsequently reached a consensus for each section. The opportunity for public comment was provided during the meeting. Following the meeting, the Panel's written evaluations, conclusions, and recommendations were consolidated as the July 2000 Peer Review Panel Report, which follows as Section I. The Panel agreed the Primary and Limit tests would perform as good or better than the existing conventional LD50 and limit tests, respectively. They also agreed the revised tests would reduce animal use compared to the current test methods. The Panel made several recommendations for revision of the UDP test guideline. The Panel did not recommend the UDP Supplemental Test.

After the July meeting, the UDP Technical Task Force prepared a revised draft test guideline (**Appendix C-1**) which incorporated and addressed the Panel's recommendations. A user friendly software program was added to aid in sequential dose selection, test-stopping decisions, calculation of an estimated LD50, and calculation of a confidence interval around the LD50. Availability of the revised draft UDP guideline and software program, and request for public comment were announced in a June 22, 2001, *Federal Register* notice (Vol. 66, No. 121, 33551-33552). A subsequent *Federal Register* notice (Vol. 66, No. 133, 36294-36295, July 11, 2001) announced the Panel meeting and requested public comment.

The UDP Panel met on August 21, 2001, via teleconference, with public meeting access made available in Research Triangle Park, North Carolina. Opportunity for public comment was provided during the meeting. The Panel reviewed and endorsed the revised UDP guideline, confidence interval procedure, and software program, with the provision that some additional clarifications should be incorporated. The Panel's evaluations, conclusions, and recommendations were consolidated as the August 2001 Peer Review Panel Report, which follows as Section II.

Following the August 2001 peer review panel meeting, the UDP Technical Task Force revised the UDP Guideline in response to the Panel's recommendations. This revised Guideline was reviewed and endorsed by the ATWG and the ICCVAM, and is provided as **Appendix B** in this report. In accordance with the ICCVAM Authorization Act of 2000, Public Law 106-545, the ICCVAM developed and adopted an ICCVAM test recommendation for the UDP, which is included in this report as **Appendix A**.

As required by P. L. 106-545, the ICCVAM test recommendation will be forwarded to Federal agencies for their consideration and appropriate actions. This publication and many of the supporting documents are available on the Internet at the ICCVAM/NICEATM website (http://iccvam.niehs.nih.gov). Agency responses to ICCVAM test recommendations will also be made available at this website.

We gratefully acknowledge all of the individuals who served as Peer Review Panel members for their thoughtful evaluations and unselfish contributions of their time. We extend a special thanks to Drs. Diane Gerken and Curtis Klaassen for their service as Panel Co-chairs, and to Drs. George Alexeeff, Wallace Hayes, Janice Kuhn, and Robert Scala for their service as Section Leaders. The efforts of the ATWG were instrumental in assuring a meaningful and comprehensive review which addressed regulatory needs. The UDP Technical Task Force was responsive to the requests and suggestions for revisions and supporting documentation over the duration of this project. The efforts of the NICEATM staff in coordinating local arrangements, providing timely distribution of information, and preparing this final report are acknowledged and appreciated. We especially thank Mr. Brad Blackard for coordinating communications and logistics throughout the entire project. On behalf of the ICCVAM, we extend our thanks to the many individuals who contributed to the evaluation of the UDP.

William S. Stokes, D.V.M.
Co-Chair, ICCVAM, NIEHS

Richard N. Hill, M.D., Ph.D.
Co-Chair, ICCVAM, U.S. EPA

Executive Summary

I. Introduction

Historical Background

The acute oral toxicity test in rodents is a critical step in defining the toxicity of a test material for the purpose of hazard classification and labeling. The acute oral toxicity test is designed to determine adverse effects and to estimate the dose level that is expected to kill 50% of the test population (i.e., the LD50).

A procedure for calculating the oral LD50 was first described by Trevan in 1927. This procedure has been used as a benchmark for comparing the toxicity of chemicals. The original method often used 50 animals or more. In 1981, the Organisation for Economic Co-operation and Development (OECD) adopted a test guideline (TG 401) for acute oral toxicity that estimated the LD50, and in some cases, the slope and confidence interval (CI). OECD TG 401 has become the traditional acute oral toxicity test. The test guideline was revised in 1987 to utilize three dose groups of five rats of one sex, with confirmation in the other sex using one group of five rats. In the absence of a range-finding study, this revision reduced the minimum number of animals used in the traditional acute oral toxicity test from 30 to 20.

In a continuing attempt to improve the estimate of acute toxicity while reducing the number of animals used per test, three alternative test methods were developed and implemented as additional OECD Guidelines for acute toxicity. These three tests are the Fixed Dose Procedure (FDP, TG 420), the Acute Toxic Class Method (ATCM, TG 423), and the Up-and-Down Procedure (UDP, TG 425).

U.S. EPA Request for Review of a Revised UDP

The U.S. Environmental Protection Agency (EPA) requested the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) organize an independent scientific peer review evaluation of the validation status of a revised Up-and-Down Procedure (UDP). The U.S. EPA forwarded the proposed "Acute Oral Toxicity: Modified Up-and-Down Procedure (Revised UDP)" to ICCVAM in April 2000. An independent peer review panel (Panel) was convened on July 25, 2000 to evaluate the proposed tests based on ICCVAM validation and regulatory acceptance criteria (NIEHS, 1997). An earlier version of the UDP test method had been adopted by the OECD TG Program in 1998 (TG 425). The revised UDP was proposed as an alternative to the existing conventional LD50 test (OECD TG 401, 1987; U.S. EPA 870.1100, 1998) used to assess the acute oral toxicity of chemicals. The U.S. EPA subsequently determined it was necessary to revise the UDP. The revisions were needed to 1) conform to a newly harmonised global hazard classification scheme for acute toxicity (OECD, 1998b; updated OECD, 2001); and 2) to incorporate changes to ensure the regulatory and testing needs would be met using the revised UDP prior to the OECD's proposed deletion of the TG 401 (OECD, 1987).

Components of the Revised UDP Test Method

The revised UDP test method submitted to ICCVAM in April 2000 included three components:

a) Primary Test, which provided an improved estimate of acute oral toxicity with a reduction in the number of animals used when compared to TG 401 and the existing TG 425;
b) Limit Test for substances anticipated to have minimal toxicity; and
c) Supplemental Test to determine the slope and confidence interval (CI) for the dose-response curve.

The Panel congratulates the agencies of the United States and the OECD for moving forward with the sequential testing of animals, as was achieved with the adoption of OECD TG 425 and in the proposed revision. Also, the development

team for the revised UDP demonstrated a comprehensive understanding of the statistical issues involved and is to be commended for the effort that went into revising the UDP Guideline.

In the revised UDP Primary Test, one animal is orally administered an appropriate dose (with 175 mg/kg as the default starting dose) and observed for up to 14 days. If the animal is alive at 48 hours after treatment, a second animal is orally administered a preset higher dose (0.5 log spacing by default). If the first animal dies, then the second animal is dosed at a preset lower dose (0.5 log spacing by default). Dosing stops when one of three stopping criteria is satisfied, with as few as six, but not more than 15 animals used per test.

In the revised UDP Limit Test, one animal is dosed at the limit dose (2000 or 5000 mg/kg). If the animal dies, the UDP Primary Test is conducted. If the animal lives, two more animals are dosed concurrently at the limit dose. If both of these animals live (i.e., three animals have survived), the UDP Limit Test is stopped. If one or both of the two animals die, additional animals are dosed sequentially at the limit dose until either three animals have survived or three animals have died (i.e., the maximum number of animals tested is five). If three animals survive, the LD50 is above the limit dose. Conversely, if three animals die, the LD50 is below the limit dose level.

In the UDP Supplemental Test for determining the slope and CI, three treatment schedules at increasing dose levels are initiated, each at a dose level that is a factor of 10- to 30-fold below the estimated LD50 obtained in the UDP Primary Test. Dosing continues in each sequence until an animal dies. All data, including data obtained in the UDP Primary Test, are then considered in a statistical model that estimates the slope and CI.

II. ICCVAM Independent Scientific Peer Review, July 25, 2000 Peer Review Meeting

In a public session on July 25, 2000, an international independent scientific peer review panel (Panel) met to evaluate the validation status of the revised UDP (*Federal Register*, NIEHS, 2000a, 2000b). The Panel was charged with evaluating the extent to which established ICCVAM validation and acceptance criteria had been addressed, and subsequently developing conclusions regarding the usefulness and limitations of the UDP. Evaluation of the Revised UDP was divided into four sections:

1. General Considerations for the Revised UDP Protocol;
2. Revised UDP Primary Test;
3. Revised UDP Limit Test; and
4. UDP Supplemental Test.

The Panel was also asked to respond to the following questions for each of the three tests:

- Has the revised UDP been evaluated sufficiently, and is its performance satisfactory to support its adoption as a substitute for the currently accepted UDP (OECD TG 425), and as a substitute for the conventional LD50 test for acute oral toxicity (U.S. EPA OPPTS 870.1100; OECD TG 401)?

- With respect to animal welfare, does the revised UDP adequately consider and incorporate where scientifically feasible, procedures to refine, reduce, and/or replace animal use?

In response to these questions, the Panel concluded the following:

1. The performance of the revised UDP Primary Test is satisfactory and exceeds the performance of OECD TG 401 in providing, with fewer animals, both an improved estimate of the LD50 for the purpose of hazard classification and more accurate information on acute toxicity. In particular, the use of 0.5 log units for dose spacing is reasonable and appropriate based on experience and the results of computer simulations. Three disadvantages of the revised UDP Primary Test recognized by the Panel were: a) the increased length of time needed to conduct a study; b) the increased costs per test material evaluated; and c) the increased complexity of the protocol.
2. The revised UDP Limit Test at 2000 or 5000 mg/kg is expected to perform as well as or

better than the Limit Test in OECD TG 401, with a reduction in the number of animals needed to conduct a test.

3. The UDP Supplemental Test for slope and CI was not recommended for adoption. The Panel was unable to evaluate the utility of the test because sufficient information regarding the use of the resulting data was not provided. As a consequence, any impact on animal use was not assessed.
4. The revised UDP Primary Test and the revised UDP Limit Test will reduce the number of animals used, but will not replace the use of animals. The Panel could not reach a consensus on the overall issue of refinement. However, the OECD Guidance Document on the Recognition, Assessment, and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation (OECD, 2000a), referenced in the revised UDP Guideline, provides an element of refinement.

The recommendations of the Panel for the revised UDP follow. Additional information can be found in the appropriate sections of this report.

General Considerations

With regard to general protocol and UDP Guideline-related issues, the Panel recommended the following:

- The use of either sex (all males or all females) should be permitted unless information is available suggesting that one sex is more sensitive;
- The use of constant volume or constant concentration of the test material during administration should be allowed;
- All reference to littermates should be excluded from the U.S. EPA Revised UDP Guideline;
- Animals of 8 to 12 weeks of age should be used;
- Individual animal body weights on the day of dosing should be within 20% of the mean body weight for all animals dosed throughout the study;
- Additional guidance detailing how to use all pre-start data (e.g., *in vitro* test results, physical and chemical properties) should be provided in the Guideline;
- The overall usefulness of information (e.g., clinical signs, time course of effects, target organs, pathology, etc.) gained beyond the LD50 should be emphasized in the UDP Guideline; and
- The Guideline should be reorganized to improve clarity.

UDP Primary Test

With regard to the revised UDP Primary Test, the Panel recommended the following:

- The scientific basis should be presented in the Revised UDP Guideline;
- Guidance for when to use the UDP Primary Test should be included in the Guideline;
- Additional guidance on the starting rule and a justification of the default starting dose of 175 mg/kg should be discussed in the Guideline;
- An improved description of stopping rule #3 should be included in the Guideline;
- User-friendly, validated software for test use or access to such software should be provided; and
- A practicability evaluation should be conducted (an appropriate working group should consider the design of this evaluation).

UDP Limit Test

With regard to the revised UDP Limit Test, the Panel recommended:

- The scientific basis and rationale should be added to the Revised UDP Guideline; and
- Additional discussion of how and where the revised UDP Limit Test is integrated into the strategy of hazard or safety assessment should be included in the Guideline (a flow chart with decision criteria covering the complete testing scheme might be an efficient way to attain this goal).

UDP Supplemental Test

With regard to the UDP Supplemental Test, the Panel recommended:

- a more clearly defined purpose of how the slope and CI are used for human and environmental risk assessment should be included in the Revised UDP Guideline; and
- Consideration should be given as to whether the slope and CI are the most appropriate parameters for risk assessment or whether risk assessment needs can be addressed more

directly. For example, if estimates of points on the dose-response curve well below the median lethal dose are needed in environmental risk assessment, more efficient methods should be considered.

Revisions to the UDP in response to the July 25, 2000 Panel Report

Based on the Panel's conclusions and recommendations from July 25, 2000, the UDP Technical Task Force revised the UDP test method guideline as follows:

- Revisions recommended by the Panel were incorporated into the proposed UDP Primary and Limit Tests;
- The UDP Supplemental Test to determine the slope of the dose-response curve was deleted;
- A procedure was added (for use with the Primary Test) to calculate the confidence interval (CI) for the estimated LD50. This procedure is a statistical calculation that does not require the use of additional animals. The CI helps to place the estimated LD50 in a statistical context for hazard and risk assessment purposes.
- The U.S. EPA developed a software program for use in establishing test doses, determining when to stop the test, estimating the LD50, and providing a CI for the LD50. The publicly available software was developed to mitigate complexity for the user and to facilitate correct performance of the UDP.

The UDP Technical Task Force provided the following clarifications regarding animal welfare:

- The UDP guideline significantly reduces the number of animals used in comparison to OECD TG 401 by the incorporation of the following: 1) a stopping rule which limits the maximum number of animals in a test; and 2) a sequential dosing method which introduces further efficiencies in animal use.
- The UDP guideline provision that the initial starting dose should be below the LD50 will result in fewer animals receiving lethal doses, thereby providing further potential reduction in pain and distress.
- Adherence to the OECD Guidance Document on Humane Endpoints (2000a) should provide additional reduction or minimization of pain and distress in animals used in this procedure.

The revised version of the UDP and the UDP software program were then provided to the Panel and made available for public comment in July 2001(*Federal Register*, NIEHS, 2001a).

August 21, 2001 Peer Review Panel Meeting

The UDP Panel met, via public teleconference, on August 21, 2001 (*Federal Register*, NIEHS, 2001b). The agenda topic of the teleconference meeting was the scientific peer review evaluation of the following:

1. The revised draft UDP, modified in response to recommendations from the July 2000 Panel meeting;
2. A proposed procedure for calculating the confidence interval (CI) for the estimated LD50; and
3. A software program to aid in establishing test doses, determining when to stop the test, estimating the LD50, and providing a CI for the LD50.

The Panel was to evaluate the following:

1. The extent to which the revised draft UDP test guideline (July 12, 2001) incorporates modifications in accordance with the Panel's recommendations at the July 25, 2000 Peer Review Panel meeting;
2. The appropriateness and adequacy of the proposed procedure for calculating a CI for the LD50; and
3. The adequacy and consistency of the software program for use in the revised draft UDP test guideline.

Conclusions and recommendations from the Panel were as follows:

Revisions to the UDP Test Guideline

The Panel concluded many of the recommended and requested changes had been appropriately considered and all members concurred with the current modifications. However, several previous recommendations appeared to have not been

adequately addressed in the revised UDP Test Guideline, and the Panel recommended adding the following:

1. Either sex of animal can be used, or if information is available indicating that one sex is more sensitive, the more sensitive sex should be used.
2. A practicability evaluation of the usability of the *in vivo* test should be conducted to supplement the computational analyses.
3. A separate section on how the revised UDP Primary Test addresses reduction, refinement, and replacement of animals when compared to the previous tests should be included to the UDP guideline.
4. Constant concentration in dosing should be used unless there is a clear scientific or regulatory justification for using constant volume. In the event that constant volume is used, information on the actual concentrations utilized should be provided.
5. Additional guidance pertaining to the use of pre-start data (data available before the acute toxicity test is conducted) which may be helpful in determining the starting dose level should be provided.

Confidence Interval Procedure

Calculation of confidence intervals (CI) provides a basis for evaluating how to incorporate test results into regulatory applications. Therefore, a CI calculation was included in previous versions of the UDP guideline (OECD 1998 and ASTM 1998). Following deletion of the proposed supplemental procedure from the previous draft Revised UDP as per recommendation by the July 2000 Panel review, another method was needed to assist the investigator using the UDP to calculate a CI for the LD50. Based on this need, the U.S. EPA developed a proposed procedure for obtaining the CI; this procedure is a statistical calculation that does not require the use of test animals beyond what is needed to estimate the LD50. Further, the procedure helps to place the estimated LD50 in a statistical context for hazard and risk assessment purposes.

The Panel endorsed the proposed procedure for calculating the confidence interval for the estimated LD50. However, the Panel recommended the inclusion of language in the UDP guideline and software to fully describe the limitations and uncertainties of the proposed method, and to provide appropriate cautions for interpretation of test results. The Panel noted that statistical techniques are evolving and recommended the future development of alternative approaches, such as nonparametric methods, be encouraged.

Software Program

To support the modifications in the revised draft test guideline, a software program was designed and made publicly available to aid in the guideline procedures, to facilitate performance of the UDP, and to mitigate its complexity for the user. The U.S. EPA developed the Acute Oral Toxicity (U.S. EPA Revised Test Guideline 425) Statistical Program" (AOT425StatPgm) to perform the statistical calculations associated with the guideline. The AOT425StatPgm program performs the calculations required to complete the test procedure by calculating 1) the doses for the test animals, 2) when to stop dosing animals, and 3) the specified LD50 and a confidence interval for the LD50. Additionally, U.S. EPA conducted quality assurance testing and simulation testing to assess the performance of the software program and to determine the statistical performance of the OECD TG 425 procedure under various conditions.

The Panel concluded the software program was appropriate and suitable for establishing test doses, determining when to stop the test, estimating the LD50, and providing a CI for the LD50.

Up-and-Down Procedure (UDP)

Peer Panel Report

July 25, 2000 Meeting

1.0 INTRODUCTION

This report summarizes the results of the July 25, 2000 independent scientific peer review panel evaluation of the revised Up-and-Down Procedure (UDP), a method proposed as a substitute for the existing LD50 test for assessing the acute oral toxicity potential of chemicals. The meeting was organized by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM), and sponsored by the National Institute of Environmental Health Science (NIEHS) and the NTP. The Peer Review Panel evaluated the usefulness of the UDP as an alternative to the conventional LD50 test method for acute oral toxicity currently accepted by regulatory authorities. *Federal Register* notices relevant to the meeting include a Request for Data and Nomination of Expert Scientists (NIEHS, 2000a) and Notice of Peer Review Meeting and Request for Comments (NIEHS, 2000b). These notices are provided in **Appendix D**.

This introduction briefly summarizes the purpose and history of acute toxicity testing and the purpose and conduct of the July 25, 2000 meeting. The remaining parts of this section summarize the UDP Peer Panel's discussions, conclusions, and recommendations from the July 25, 2000 meeting. A report on a follow-up meeting of the peer review panel on August 17, 2001 is provided in Section II. **Appendix A** provides ICCVAM Test Method Recommendations on the UDP, **Appendix B** contains the Final Revised U.S. EPA UDP Test Guideline which addresses the recommendations from both Panel, **Appendix C** contains the materials reviewed by the Panel for the August 2001 Peer Panel Meeting, and **Appendix E** provides Summary Minutes and Public Comments from the UDP meetings. **Appendix F** provides the Background Review Document on the UDP which has been revised to incorporate many of the recommendations and suggestions from the Panel at the July 2000 meeting. **Appendices G** through **P** provide additional background information about the UDP Primary Test, Limit Test, and Supplemental Test which was reviewed by the Panel in preparation for their July 2000 meeting. **Appendix Q** summarizes the relevant U.S. Federal Regulations on Acute Oral Toxicity.

1.1 History and Purpose of Acute Toxicity Testing

Acute oral toxicity testing is conducted to determine the hazard potential of a single oral exposure to various chemicals and products. The acute oral toxicity test in rodents is a critical step in defining the toxicity of a test material for the purpose of hazard classification and labeling. It is designed to determine adverse effects and to estimate the dose that is expected to kill 50% of the test population (i.e., the LD50).

Four regulatory agencies in the United States, the Department of Transportation (DOT), the Consumer Product Safety Commission (CPSC), the Occupational Safety and Health Administration (OSHA), and the U.S. Environmental Protection Agency (EPA) require industry to label chemicals and products with hazard information based on LD50 estimates. DOT requires oral lethality data to determine the transportation requirements for hazardous substances (49 CFR 173). CPSC requires such information for labeling hazardous substances so as to protect consumers when such products are used in the home, the school, and recreational facilities (16 CFR 1500). OSHA requires the use of acute lethality data to implement labeling requirements for the hazard communication program to protect employees (29 CFR 1910). Certain U.S. EPA regulatory programs also require the submission or generation of acute toxicity data for hazard classification purposes (40 CFR 156). During acute toxicity testing, non-lethal endpoints may also be evaluated to identify potential target organ toxicity, toxicokinetic parameters, and/or dose-response relationships.

As shown in Table 1, the international community also uses acute oral toxicity data as the basis for hazard classification and the labeling of chemicals for their manufacture, transport, and use (OECD, 1998b; updated OECD, 2001). Other potential uses for acute toxicity testing data include:

- Establishing dosing levels for repeated-dose toxicity studies;
- Generating information on the specific organs affected;
- Providing information related to the mode of toxic action;
- Aiding in the diagnosis and treatment of toxic reactions;
- Providing information for comparison of toxicity and dose response among substances in a specific chemical or product class;
- Aiding in the standardization of biological products;
- Aiding in judging the consequences of single, high accidental exposures in the workplace, home, or from accidental release;
- Serving as a standard for evaluating alternatives to animal tests.

Table 1.1 Adapted from the Harmonised Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixtures: Acute toxicity hazard categories and (approximate) LD50/LC50 values defining the respective categories (OECD 1998b; updated OECD, 2001)

Acute Toxicity Route	Toxicity Class 1	Toxicity Class 2	Toxicity Class 3	Toxicity Class 4	Toxicity Class 5
Oral LD50 Values (mg/kg) [approximate]	$\leq$5	>5 $\leq$50	>50 $\leq$300	>300 $\leq$2000	>2000 $\leq$5000

Historically, lethality has been the primary toxicological endpoint in acute toxicity tests. Trevan (1927) was the first to attempt to standardize a method for assessing the toxicity of potent biological toxicants, the progenitor of the "lethal dose, 50% (LD50) test". The classical LD50 test procedure evolving from this innovation in the 1970s and early 1980s used from 100 to 200 animals per test substance (Galson, 2000). Although other information, such as the slope of the dose-response curve, confidence interval for the LD50, and toxic signs, could also be obtained from this test, the procedure was severely criticized for both scientific and animal welfare reasons (Zbinden and Flury-Roversi, 1981). These criticisms eventually resulted in the proposal and adoption of a new guideline (OECD TG 401; OECD, 1987) which utilized three dose groups of five rats of one sex, with confirmation in the other sex using one group of five rats. In the absence of a range-finding study, this revision reduced the minimum number of animals used in the traditional acute oral toxicity test from 30 to 20. This method has become the most widely used for defining the acute toxicity of a chemical and a mandatory-testing requirement for new chemicals.

More recently, the acute toxicity test procedure has been modified in various ways to refine and further reduce the number of animals used to a maximum of 16 (e.g., OECD Test Guidelines 420, 423, and 425). The Globally Harmonised Scheme for Hazard Classification (OECD 1998b; updated OECD, 2001) prompted a re-assessment of all of the OECD *in vivo* test guidelines for acute toxicity (i.e., fixed dose, up-and-down procedure, acute toxic class method) to ensure that regulatory needs are met while minimizing animal usage and maximizing data quality.

Several other test designs, including the moving average (Weil, 1983), acute toxic class method (Schlede et al., 1994), and UDP (Bruce, 1985),

have been proposed. The classical experimental method for estimating the LD50 was to orally dose individual animals, in groups of five or ten per sex, with varying concentrations of the test material and to observe whether the animal lived or died over a defined period of time (generally 14 days). The method was standardized in 1981 by the international acceptance of Test Guideline (TG) 401 (OECD, 1981).

The test material is typically administered by oral gavage to fasted young adult animals. The animals are observed periodically during the first 24 hours with special attention given to the first four hours, then at least once a day for 14 days or until they recover. Clinical signs, including time of onset, duration, severity, and reversibility of toxic manifestations, are recorded at each observation period. Body weights are determined pre-treatment, weekly thereafter, and at the death of the animals or termination of the study. All surviving animals are humanely killed at 14 days or after recovery. Gross necropsies are conducted on all study animals. Variation in the results due to inter-animal variability, intra- and inter-laboratory variability, and to differences in strain, sex, estrus cycle, and species have been characterized. Based on intra- and inter-laboratory testing, the point estimate of the LD50 appears to be reliable within a factor of two or three (Griffith, 1964; Weil et al., 1966; 1967).

Although the experimental method as to dosing, handling, and observing the animals has not varied, many attempts have been made to reduce the number of animals used while maintaining the accuracy of the method for estimating the LD50. These changes in sampling technique do not involve a change in the actual treatment of the animals or in the endpoints examined.

1.2 Objectives of the July 25, 2000 Meeting

The meeting was convened to conduct an independent scientific peer review evaluation of the validation status of the revised UDP. This procedure is an updated version of the OECD Test Guideline 425 (OECD, 1998a). The revised UDP was proposed as a substitute for the existing OECD Test Guideline 401 (OECD, 1987). OECD has proposed that Guideline 401 should be deleted since three alternative methods are now available. Prior to deletion of Guideline 401, U.S. agencies requested that ICCVAM conduct an independent peer review of the revised UDP to determine the validity of the method as a substitute for Guideline 401. The Independent Peer Review Panel was to (1) evaluate the extent to which established validation and acceptance criteria (ICCVAM, 1997) have been addressed, and (2) to provide conclusions and recommendations regarding the usefulness and limitations of the method as a substitute for the traditional acute oral toxicity test method (OECD, 1987). The UDP has the potential to reduce the number of animals required to classify chemicals for acute oral toxicity compared to Guideline 401.

1.3 Conduct of the Meeting and Reports

The UDP Peer Panel Review Meeting, which was open to the public, was conducted on July 25, 2000. The meeting began with an introduction including an overview of the peer panel review process and a summary of current Federal agency requirements. The Panel then discussed the Revised UDP Protocol, Primary Test, Limit Test, and Supplemental Test. Following the final public comment session, the Panel provided conclusions and adjourned. Following the meeting the Panel prepared this written report summarizing their discussions, conclusions, and recommendations.

In this Panel report, all references made to the background review document (BRD) refer to the April 2000 BRD which can be found at http://iccvam.niehs.nih.gov/methods/udpdocs/AllBRDlk.pdf. The April 2000 BRD was revised in response to recommendations of the Panel and this revised version has been provided in **Appendix F**. When possible, both the former (April 2000) and the current reference (October 2001) for appendices and other documentation have been provided.

2.0 GENERAL CONSIDERATIONS

A laboratory-based, practical viewpoint was taken in evaluating the U.S. EPA Revised UDP Guideline (April 2000; formerly **Appendix C**, currently **Appendix G**). Consideration was given as to whether the procedures were described unambiguously, were workable in the laboratory setting, and comprised a sound basis for obtaining the necessary acute oral toxicity information without undue increases in time and expense.

2.1 Revised UDP Protocol

The type of information on the test material that should be obtained and considered prior to conducting a study is appropriately described. In general, guidance concerning the selection of the appropriate species, strain, and age of animal for testing is sufficient and appropriate. However, the revised Guideline contains an impractical reference to assigning littermates randomly to test groups. At animal receipt, the laboratory does not know which animals are littermates. In addition, since the total number of animals that will be used during a study cannot be predicted, at least fifteen animals must be assigned prior to study start. Because animal use is sequential, the study design itself minimizes bias.

Unless information is available indicating that one sex is more sensitive than the other, the use of either all males or all females should be considered to allow for additional flexibility and to decrease the total number of animals that are purpose-bred for acute oral toxicity testing. Data provided in the Background Review Document (BRD) (formerly EPA Document 14, Part A, Table 1, currently **Appendix P-1**, Table 1 on page P-6) suggest, in general, a low incidence of studies with a sex-related effect. However, gender-dependent differences in xenobiotic metabolism are more pronounced in rats when compared to other rodent species. The differences primarily involve cytochrome P450s (CYP), sulfotransferases, glutathione transferases, and glucuronyl transferases (Mulder, 1986; Nelson et al., 1996). Studies of chemicals with known sex-related differences in toxicity, attributable to differences in metabolism, have shown that females are often more susceptible when compared to males (see former U.S EPA Document 14 in the BRD, currently **Appendix P**).

Descriptions of the accepted weight range and procedures for minimizing weight variation during the test procedure are not adequate. The age and weight ranges are not specified in the April 2000 revised Guideline (formerly **Appendix C**, currently **Appendix G**) as they are in OPPTS 870.1100, which requires rats to be between eight and 12 weeks of age at the time of dosing. In addition, individual body weights recorded on the day of dosing must be within 20% of the mean body weight for all animals dosed during the study. Similar guidance is recommended in the revised Guideline.

Guidance regarding procedures for preparing animals for study and the description of dose preparation procedures is sufficient and appropriate. Guidance regarding dose administration, including dose volumes and stability considerations (e.g., the need for appropriate stability data if a single dosing solution is used over several days) should be further refined in the U.S. EPA Revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**). The use of constant concentration (in addition to constant volume) should be included as an option for at least some types of test materials. OPPTS 870.1100 requires liquids to be administered neat or at the most concentrated workable dilution, if dilution of a liquid or suspension of a solid is needed. This issue may be important in particular when testing at the limit dose (i.e., 2000 or 5000 mg/kg) to simulate accidental exposure to the undiluted product.

The notion that the test material concentration in dosing solutions might need to be supported by analytical analysis is especially burdensome, as it would greatly increase the cost. The use of constant volume dosing solutions instead of constant concentration solutions would potentially increase the analytical task and is not recommended. The cost of analytical analysis may impact the willingness of some laboratories to use the revised UDP. OPPTS does not require analytical evaluation. If it is suspected that the test material is unstable in solution, a fresh

mixture should be prepared prior to each administration. The absence of a concurrent vehicle control is justified sufficiently.

Paragraph 27 of the Revised Guideline (formerly **Appendix C**, currently **Appendix G**) provides an adequate description of appropriate observations to be recorded. The reference to Chan and Hayes (Chapter 16. Acute Toxicity and Eye Irritancy. *Principles and Methods of Toxicology*. Third Edition. A.W. Hayes, Editor. Raven Press, Ltd., New York, USA, 1994) should be removed. It may be more appropriate to include specific references in a guidance document. The first two sections of paragraphs 26 and 27 of the revised Guideline (April 2000) are repetitive and contradictory. We recommend replacement of the first sentence in paragraph 26 with the first sentence of paragraph 27. Each time the 48-hour observation interval is mentioned, as in "each animal should be observed carefully for 48 hours (unless the animal dies)", the qualifier "but need not be rigidly fixed" should be added as delayed mortality will occur often. Also, "time of death" should be worded as "time found dead" as it is unlikely the exact time of death will be determined, unless a moribund kill has been conducted.

Appropriate endpoint(s) for humanely killing animals prior to the end of the required holding period are sufficiently and appropriately described. Frequency of body weight measurements and procedures for pathology evaluations are described appropriately**.**

The description of the data to be collected and reported is largely standard guideline wording and is sufficient as such. A specific rationale for the starting dose and dose progression should be provided only when it varies from the standard described in the revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**), and removal of the requirement for justification of starting dose and dose progression when the defaults are used is suggested. However, one Panel member suggested that a rationale be provided for all starting doses and dose progressions even when the default is used. It would be helpful if a table of log doses from 0.1 log to 0.5 log was provided, starting at 10 mg/kg and progressing to 5000 mg/kg.

Procedures for recording and storing data, including suggested forms or formats, are described sufficiently. Descriptions of equipment, materials, and supplies needed are appropriate. However, a comprehensive, validated software package should be developed and distributed to assist in conducting all variations of the UDP protocol. Ideally, a series of data sets (testing program) should be provided for the purpose of "in-house" validation for compliance with Good Laboratory Practice (GLP) guidelines.

2.2 Animal Welfare Considerations (Refinement, Reduction, Replacement)

With regard to the Revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**), the majority of the Panel concluded that the validation studies and simulations appear to have demonstrated that the number of animals necessary for the revised UDP Primary Test (i.e., between six and 15) and the revised UDP Limit Test (between three and five) are appropriate to obtain scientifically valid results**.** However, some Panel members were concerned that the optimal numbers of animals for each test had not been adequately demonstrated.

The majority of the Panel concluded that the procedures in the revised UDP addressed the potential for pain and distress issues based on the inclusion of the OECD Guidance Document on the Recognition, Assessment, and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation (OECD, 2000a; formerly **Appendix B**, but no longer appended in this final report). However, the Panel concluded that only limited or no improvement was made in the area of replacement, especially for the UDP Supplemental Test. The Panel felt that additional information would be needed to adequately evaluate the UDP Supplemental Test.

The rationale for the necessity to use animals to determine acute oral toxicity is appropriate and justified, although there is an implication that the reason for not testing in humans is a legal issue rather than a moral one. The revised UDP

Guideline (formerly **Appendix C**, currently **Appendix G**) states that the primary reason for conducting animal tests is for the protection of humans from the consequences of exposure to unsafe products. However, product testing also benefits wildlife, domesticated animal, and pets.

2.3 Other Considerations

The procedures for the observation and reporting of clinical signs are appropriate and adequate for regulatory needs. However, the procedures for considering delayed deaths need clarification.

Based on the revised Guideline and the supporting documentation, the proposed test methods can be readily conducted in GLP-compliant laboratories. The procedures take more time and are more cumbersome than OECD TG 401 (formerly **Appendix A**, currently **Appendix I**) or OPPTS 870.1100. Explanation of the statistics in the revised UDP Primary Test and the UDP Supplemental Test accompanied by illustrative examples (perhaps in the form of flow charts in an appendix to the April 2000 Guideline) will be critical for the non-statistician to conduct these studies. As mentioned previously, a comprehensive, validated software package should be made available to assist with these calculations.

A reordering of the presentation of the three different types of studies in the revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**) is recommended. The revised UDP Limit Test should be described first. Additional guidance should be included to provide for a transition from the revised UDP Limit Test to the revised UDP Primary Test, when necessary.

Personnel training and experience requirements are adequately described and reasonable. The necessary equipment, materials, and supplies (e.g., animals, and computers) should be readily obtainable.

The estimated cost of an UDP study provided in the April 2000 BRD is not realistic. The cost of conducting the revised UDP Primary Test will be greater than the traditional acute toxicity test, perhaps up to twice as much, due to the needs for increased technical expertise, specialized statistical analysis, as well as to the difficulty associated with scheduling (animal shipments, dose preparation, dosing, necropsy) and organizing the data for reporting. For example, the challenge of scheduling multiple simultaneous UDP Primary Tests is much greater than that associated with the scheduling of the same number of OECD TG 401 (formerly **Appendix A**, currently **Appendix I**) tests. Ensuring that adequate numbers of animals in the appropriate weight range are readily available will be more difficult than would be for the traditional LD50 test. Laboratories that infrequently conduct the UDP test may be forced to humanely kill a greater number of undosed animals. As a consequence, particularly for smaller companies with limited resources, the difference in product testing costs could be significant.

Depending on study progression, it is likely that the revised UDP Primary Test will take significantly more time than the traditional acute toxicity test. Realistically speaking, it is difficult to dose more than two animals per week unless one of the treated animals dies on treatment day. If dose levels are started close to the LD50, animals generally take two to three days to show morbidity/mortality. Therefore, the revised UDP Primary Test will most likely take at least three weeks if the minimal number of animals (i.e., 6) is used and seven to eight weeks if the maximum number of animals (i.e., 15) is used. Although not recommended by the Panel, addition of the UDP Supplemental Test would increase the total duration of the study by an additional two to five weeks per test material. In contrast, the traditional acute toxicity test using three dose levels generally takes four to five weeks and yields a similar amount of information.

In reference to the revised Guideline (formerly **Appendix C**, currently **Appendix G**), the outcome of the UDP Primary Test is likely to be sensitive to differences in dose selection and progression as well as to the statistical procedures employed. This revised UDP Primary Test protocol has now become even more complicated than the current UDP (OECD, 1998; former **Appendix A**, current **Appendix H**) and the results are probably very sensitive to errors in dose level

selection. The more complicated the protocol, the more extensive the measures that must be taken to minimize the likelihood of errors in the laboratory.

2.4 Recommendations

1. The U.S. EPA Revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**) should be re-ordered to present the revised UDP Limit Test first since this test is more likely to be used for the majority of test materials.
2. Additional guidance on the transition from the revised UDP Limit Test to the revised UDP Primary Test, when appropriate, should be provided in the revised Guideline.
3. All reference to littermates should be excluded from the revised UDP Guideline (April 2000; formerly **Appendix C**, currently **Appendix G**).
4. The use of either sex (all males or all females) in a study should be allowed unless information is available suggesting that one sex is more sensitive.
5. The use of animals of 8 to 12 weeks of age at the time of dosing should be specified in the revised Guideline.
6. The revised Guideline should state that individual animal body weights on the day of dosing must be within 20% of the mean body weight for all animals dosed.
7. The option for constant concentration in addition to constant volume solutions should be included in the revised Guideline.
8. In the U.S. EPA Revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**), the Chan and Hayes (1994) reference and the first sentence in paragraph 26 should be deleted. Paragraph 27 provides an adequate description of the clinical observations to be conducted. In addition, the qualifier of "but need not be rigidly fixed" should be added to "48 hours".
9. A table of log doses from 0.1 log to 0.5 log, starting at 10 mg/kg and progressing to 5000 mg/kg, should be included in the revised Guideline.
10. A comprehensive, validated software package should be developed and distributed to assist in conducting all variations of the UDP protocol. Ideally, a series of data sets (testing program) should be provided for the purpose of "in-house" validation for compliance with GLP guidelines.

3.0 REVISED UDP PRIMARY TEST

3.1 Introduction and Rationale for the Revised UDP Primary Test

3.1.1 *Scientific Basis for the UDP Primary Test*

Inadequate information on the **scientific** basis of the revised UDP Primary Test (e.g., what information is needed about acute toxicity, how the test results would be used) was provided in the U.S. EPA Revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**) and in the April 2000 BRD. The **technical** basis for the revised UDP Primary Test is described in detail; however, the description is not completely understandable and requires clarification. Paragraph 10 of revised UDP Guideline [Principle of the Primary (Single Estimate) Test] and the corresponding Section 1.2 the April 2000 BRD (The Scientific Basis of Revised UDP) appear to discuss different issues; paragraph 10 provides a synopsis of the test method while Section 1.2 provides information about the philosophy behind the procedure. Consequently, it is difficult to reconcile the information provided in these two sections. Nonetheless, the technical basis for the revised UDP Primary Test is, for the most part, adequately described. The literature reference on page C25 of the April 2000 BRD is incomplete; for reference number 14, the date is 1994.

3.1.2 *Intended Uses of the Revised UDP Primary Test*

In the revised Guideline (formerly **Appendix C**, currently **Appendix G**), the rationale for the revised UDP Primary Test is clearly presented. By concentrating testing around the LD50, the UDP requires fewer animals per study than OECD TG 401 (formerly **Appendix A**, currently **Appendix I**). Should the starting dose be far from the LD50, a bias may be introduced. This bias is true particularly for test materials with a shallow slope for the dose-response curve; in addition, the bias is reduced relative to OECD TG 425 (formerly **Appendix A**, currently **Appendix H**) by the increased progression factor between consecutive doses. It is stated that the revised UDP will replace the current regulations on acute oral toxicity testing for the Consumer Product Safety Commission (CPSC), the U.S. EPA, and the U.S. Department of Transportation (DOT). However, it appears that both the U.S. EPA and the U.S. DOT already use this revised UDP Primary Test and that only the CPSC will be adopting this protocol as a new procedure. The justification provided is that the use of the revised UDP Primary Test will enhance the ability of the CPSC to use data for risk assessment purposes and for probabilistic modeling; information is not provided about the scientific basis of the test.

If the observations of animals administered a low dose demonstrate a no-observed-adverse-effect-level (NOAEL), these data may be used to estimate an acute reference dose when considering residues of highly toxic pesticides in foods. It appears that the revised UDP Primary Test (April 2000) provides a better estimate of the LD50 for classification when compared to OECD TG 401 (formerly **Appendix A**, currently **Appendix I**). A summary table comparing simulation results for the April 2000 revised UDP Primary Test with OECD TG 401 in a format similar to that on former page C-401, current page O-13 of the BRD would be helpful.

Neither the revised Guideline, the April 2000 BRD, nor the oral presentation at the July 2000 Panel meeting provided sufficient information for evaluation of how the revised UDP Primary Test will be integrated into the U.S. EPA's strategy for assessing the hazard or safety of materials. The types of materials that are amenable to the test have been delineated. The test is designed for materials that can be administered neat (without dilution) or in a solvent. The test is not restricted to materials that are water-soluble. Any solvent or vehicle can be used, but the solvent or vehicle must not add to or mask the toxicity of the test material. Although the proposal did not specifically address biopesticides, there should be little concern about testing these materials with the revised UDP Primary Test procedure. The revised Guideline stated that the LD50s of materials with shallow slopes are underestimated.

The Panel had two concerns regarding the 25 test materials used to validate the revised UDP (Bruce, 1987, Bonnyns et al., 1988, Yam et al., 1991).

First, in the Bruce (1987) validation study, eight of the 10 test materials were proprietary. As a consequence, their chemical class is unknown and some members of the Panel expressed doubt as to whether these data should have been considered for validation. Second, as each of the 25 test materials was tested in a single laboratory only, no assessment of interlaboratory reproducibility was possible. However, with the exception of mercury chloride, there was excellent concordance in the estimated LD50 between OECD TG 401 (formerly **Appendix A**, currently **Appendix I**) and the current UDP (formerly **Appendix A**, currently **Appendix H**).

3.2 Revised UDP Primary Test Protocol

A statement is made in the U.S. EPA Revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**) that all information on the material to be tested should be considered. However, no details were provided about the nature of the information to be obtained or how such information should be considered. Thus, prior to study start, a general description of the information (e.g., *in vitro* data, physicochemical properties, etc.) for consideration should be provided; in addition, how such information should be used to predict the need for the study and/or the starting dose should be determined [for example, Spielmann et al., (1999) provides information that could be useful].

A precise description of what is meant by the "slope" of the dose-response curve should be included in the Guideline. Also, in paragraph 18 of the revised Guideline (formerly **Appendix C**, currently **Appendix G**), the sentence stating, "however, when justified by specific regulatory needs, testing up to 5000 mg/kg body weight may be considered" needs to be clarified (i.e., when is it a requirement, and if not, what would justify testing at the higher limit dose?). In the revised Guideline, a "similar" dose progression should be reworded to the "same" dose progression. The April 2000 BRD (Section 1.1.5) states that the default starting dose of 175 mg/kg was chosen based on historical data and the results of computer simulations; further justification of this starting dose is needed.

The revised Guideline should include a more comprehensive description of the information needed to select an appropriate value for the slope, of when to use the default dose progression factor, and of the methods to be used in the final analysis. Because the dose progression factor can have a large effect on bias if chosen inappropriately, it should be stated that a value other than the default should be used only if there is clear evidence that the slope of the dose-response curve is far from a value of two.

The term "half-log spacing" is more accurate than a dose spacing factor of 3.2. It should be defined and used consistently throughout. The use of half-log units appears to lead to a reasonable estimate of the LD50, although no direct comparisons with other possible values were found in the simulation study results. The relatively large value reduces the bias when the starting dose is far from the true LD50 because the testing dose approaches the LD50 rapidly. This spacing allows one to reach 2000 or 5000 mg/kg with considerably fewer animals than the original 1.2 progression factor. The disadvantage is that when testing does occur near the LD50, the final estimate of the LD50 is less precise due to the larger dose spacing. An extreme example is for materials with steep slopes (above about 4); in such studies, dose levels often exhibit 100% mortality or 100% survival. The estimated LD50 is known only to occur between the lowest fatal dose and the highest non-fatal dose. This type of data occurs also in the methods described in OECD TG 420 and OECD TG 423 (formerly **Appendix A**, but not included in this final report**)**, which do not provide an estimate of the LD50. Any estimate of the LD50 resulting from the UDP depends on the choice of the assumed dose-response curve slope. A similar situation arises when both death and survival occur at a single dose level only. It would be interesting to know how often this finding was observed in the simulations.

In the revised Guideline and in the April 2000 BRD, the description of stopping rule #3 is not provided in sufficient detail and some aspects are confusing and/or scattered throughout the documents. The information could be consolidated and clarified. Terms like "the

number of animals after the first reversal" should be more clearly defined. A single software package allowing implementation of all three stopping rules should be developed and evaluated in an *in vivo* practicability study.

Computer simulation results show clearly that using the revised stopping decision criterion reduces the effect of an outlier on the estimate of the LD50 relative to the estimate obtained using OECD TG 425 (formerly **Appendix A**, currently **Appendix H**). There does not appear to be any specific evidence regarding reliability, though the reliability of the U.S. EPA Revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**) would likely be comparable to OECD TG 401 and OECD TG 425 (formerly **Appendix A**, currently **Appendices I and H**, respectively). The Guideline should be modified to allow estimation of the LD50 by any suitable statistical method (e.g., isotonic regression).

3.3 Performance of the Revised UDP Primary Test

3.3.1 *Characterization of Materials Tested*

Given that this test represents a modification of OECD TG 425 (formerly **Appendix A**, currently **Appendix H**) only, simulation studies seem to be an appropriate method of assessment. The simulation studies include materials with a full range of LD50 and slope values. However, the range of dose-response slopes is not clearly discussed in Sections 3 or 6 of the April 2000 BRD.

3.3.2 *Performance of the Revised UDP Primary Test*

The conclusions on the usefulness of the April 2000 revised UDP Primary Test are appropriate based on computer simulations. Since no formal *in vivo* validation has been reported for the revised UDP Primary Test, at a minimum, a practicability evaluation of the revised test should be conducted. The performance of the revised UDP Primary Test has been adequately described. The revised UDP Primary Test better predicts the LD50 when compared to the traditional acute toxicity test method (OECD TG 401; formerly **Appendix A**, currently **Appendix I**). However, although the revised test method uses fewer animals, the study duration in most cases will be longer. Costs for the revised UDP Primary Test and OECD TG 401 (formerly **Appendix A**, currently **Appendix I**) are reported in the April 2000 BRD to be similar, but in reality appear to be greater.

With regard to the revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**), the primary limitation of the revised UDP Primary Test is the poor estimation of the LD50 for test materials with shallow slopes for mortality. This limitation is common to all of the proposed test methods. Since only a small number of chemicals have been evaluated in the current UDP (formerly **Appendix A**, currently **Appendix H**), the extent of this limitation cannot be defined with any degree of assurance. However, according to the April 2000 BRD, it is stated that any class of chemicals or products that can be tested using OECD TG 401 (formerly **Appendix A**, currently **Appendix I**) can be tested using the revised UDP. The April 2000 BRD further states that this test method is designed for materials that can be administered neat or in a solvent. The test method is not restricted to materials that are water-soluble; any solvent or vehicle can be used as long as the solvent or vehicle does not add to or mask the toxicity of the test material. These are logical statements, but insufficient data are available to support these assertions.

3.4 Reliability (Intra-laboratory Repeat-ability; Intra- and Inter-laboratory Reproducibility) of the Revised UDP Primary Test

In the revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**), the estimated intra- and inter-laboratory reliability of the revised UDP Primary Test appears to be acceptable and better than that for OECD TG 401 (formerly **Appendix A**, currently **Appendix I**). Although the reliability is likely to be very similar to that for OECD TG 425 (1998) and even for OECD TG 401 (1987), Section 7 of the April 2000 BRD states "there are no known *in vivo* data on the reliability and repeatability of the revised UDP." In the limited testing that has been conducted, the UDP has been shown to perform

well when compared to OECD TG 401. A number of the test materials evaluated in the Bruce study (1987) were unidentified and only a small number of materials were examined in the Bonnyns et al. (1988) and Yam et al. (1991) studies, with no single material tested in more than one laboratory. Additional computer simulations should be conducted to assess the effect of changing response probabilities with the age and weight of the animals at the time of treatment.

3.5 Summary Conclusions

With regard to the revised Guideline, the revised UDP Primary Test is a suitable replacement for OECD TG 401 (formerly **Appendix A**, currently **Appendix I**). Most information obtained with OECD TG 401 is also obtained with the revised UDP Primary Test (e.g. classification, point estimate, acute toxicity characteristics). There is substantial reduction in the number of animals required, but no or little improvement in the areas of refinement or replacement.

It appears that the revised UDP Primary Test provides a better estimate of the LD50 for classification and the potential for better overall information on acute toxicity with fewer animals when compared to OECD TG 401.

3.6 Recommendations

1. The scientific basis for the test should be enhanced and added to the April 2000 Guideline, with greater explanation in the April 2000 BRD.
2. The revised Guideline should include a description of how historical data should be used to decide when to use the UDP Primary Test, the UDP Limit Test, or not to conduct any test.
3. Justification should be provided in the revised Guideline as to why the recommended starting dose of 175 mg/kg (in the absence of any relevant information) should be used.
4. In the Guideline, stopping rule #3 should be clearly defined and justified.
5. A single software package covering the entire procedure and including all three stopping rules should be developed.
6. In the U.S EPA revised Guideline, stopping rule #1 of the UDP Primary Test and the UDP Limit Test should be harmonized.
7. In the Guideline, the term “half-log” units should be used throughout rather than the approximate dose progression factor of 3.2.
8. A table of computer simulations comparing the revised UDP Primary Test with OECD TG 401 (formerly **Appendix A**, currently **Appendix I**) should be included in the BRD (e.g., see the table on page O-13 of **Appendix O-2** (former page C-401) comparing the original UDP with OECD TG 401). The simulations should include an assessment of the effect of changing response probabilities with the age and weight of the animals at the time of treatment.
9. Since no formal *in vivo* validation has been reported for the revised UDP Primary Test, at a minimum, a practicability evaluation of the revised test should be conducted.
10. The April 2000 BRD should include a separate section discussing how reduction, refinement, and replacement (i.e., the 3 R’s) are addressed by the revised UDP Primary Test.
11. In the U.S. EPA Revised UDP Guideline, the overall usefulness of information (e.g., clinical signs, time course of effects, target organs, pathology, etc.) gained beyond the LD50 in the revised UDP Primary Test should be emphasized.
12. It is recommended that either sex can be used unless information suggests one sex is more sensitive.
13. The term “slope” should be defined in the April 2000 Guideline and BRD.
14. The revised Guideline should state that any suitable statistical LD50 estimate method (e.g., isotonic regression) might be used.

4.0 REVISED UDP LIMIT TEST

4.1 Introduction and Rationale for the Revised UDP Limit Test

With regard to the U.S. EPA Revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**), the scientific basis for the revised UDP Limit Test is not adequately described in either the Guideline or the April 2000 BRD. A brief description of how to conduct the UDP Limit Test is provided, but no explanation of the scientific basis or the rationale for the revised test is reported. A scientific basis would explain why the proposed approach produces valid estimates and would provide a description of the advantages of the revised UDP Limit Test over other methods. The scientific basis should be added to the revised Guideline, with greater explanation in the BRD.

The rationale for the revised UDP Limit Test as a substitute test method for existing regulatory acute toxicity limit test methods, such as OECD TG 401 (formerly **Appendix A**, currently **Appendix I**), is not adequately described. It would be helpful to explain why the revised UDP Limit Test is a suitable replacement of the Limit Test in OECD TG 401. The rationale should describe the conclusions that could be made using the revised UDP Limit Test. The primary conclusion of the revised UDP Limit Test is that the LD50 is either above or below the limit dose used in the test. The discussion in the April 2000 BRD describes the potential uses of the revised UDP Primary Test, but not the revised UDP Limit Test. Consequently, additional discussion of the functionality of the revised UDP Limit Test in the strategy of hazard or safety assessment would significantly strengthen the revised Guideline. A flow chart with decision criteria for the entire testing scheme might be an efficient way to characterize this relationship. A chart would help also to place the revised UDP Limit Test in perspective to other tests as well as explain its relationship to the revised UDP Primary Test and any supplemental tests.

4.2 Revised UDP Limit Test Procedure

In the U.S. EPA Revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**), the procedures for conducting the revised UDP Limit Test merit further clarification. Specifically, further explanation is needed in the Guideline regarding the scientific basis, the selection of the limit dose, the stopping rule, how the revised UDP Limit Test is integrated into the revised UDP Primary Test, and factors that may set the two tests apart. These Guideline clarifications would improve the usability of the test and reduce confusion in its implementation.

While the scientific basis and rationale for the revised UDP Limit Test should be stated in the April 2000 BRD, a short statement, similar to that for the revised UDP Primary Test, would also be helpful in the revised UDP Limit Test Guideline. The revised Guideline would be improved if a short rationale such as the following were added: "Principle of the Limit Test: When it is necessary to determine if (or confirm) that the LD50 is above a defined limit (2000 or 5000 mg/kg), the UDP Limit Test may be performed." This or a similar statement would help explain the general purpose of the revised UDP Limit Test.

Clarification of the selection of the limit dose would be helpful in the April 2000 Guideline and BRD. The description of the revised UDP Limit Test specifies a limit dose of 2000 mg/kg with the option of using 5000 mg/kg. This option reflects the difference between European and U.S. testing. However, this difference is not discussed in the Guideline or the BRD and inclusion of such information would be helpful. Further, the Guideline and BRD state "dosing should not normally exceed 2000 mg/kg body weight." This statement could be interpreted in several different ways and requires greater clarity. The BRD implies that 2000 mg/kg is the standard limit dose, but in some cases 5000 mg/kg may be used. However, one section of the April 2000 BRD (Section 6.3.3.2) differs from the other sections in that it mentions a lower testable dose. Discussions indicated that in some circumstances the limit dose could be less than 2000 mg/kg. The Panel is concerned that tests with lower limit doses may be inappropriate and may confuse

standardization of guidelines. The rationale for conducting a test at a limit dose lower than 2000 mg/kg should be clearly explained in the BRD.

The stopping rules are explained in the April 2000 Guideline (Paragraph 23) and in the April 2000 BRD (Section 2.1.4). The basic stopping rule in the revised UDP Limit Test is the occurrence of two additional survivors or three deaths following survival of the first animal. This rule differs from the stopping rule that would be applied when reaching the upper bounding limit during the revised UDP Primary Test, which requires that three consecutive animals survive. The two different stopping rules may cause confusion. This issue needs to be clarified in the Guideline and the BRD.

With regard to the revised Guideline, guidance was not provided as to the next action to take when the test does not demonstrate that the LD50 is above the limit dose tested. The Guideline should state clearly that, depending on the pretest question, testing either stops or the revised UDP Primary Test should be conducted. Furthermore, in Limit Test studies in which three animals fail to survive, it should be stated explicitly that the results do not provide **any** scientifically relevant information about the actual value for the LD50. Integration of the revised UDP Limit Test into the testing strategy would clarify how the testing should be approached. As recommended previously, the revised UDP Limit Test section should precede the revised UDP Primary Test section.

The April 2000 revised UDP Limit Test, which allows the conclusion that the LD50 is greater than the limit dose if three animals, including the first, survive, is much less stringent than OECD TG 425 (in which six consecutive animals, three of each sex, must survive), but slightly more stringent than OECD TG 401 (in which at least five of ten animals must survive). In the BRD, the probability calculations (formerly EPA Document 7, **Appendix C**; currently, **Appendix M**) show that the performance of the proposed sequential method is very similar to that of a method where the number of animals tested is fixed (e.g., OECD TG 401 Limit Test; formerly **Appendix A**, currently **Appendix I**). However, the reduction in sample size results in an increased probability of misclassification for materials with an LD50 above the limit dose, especially when the LD50 is close to the limit dose. More discussion in the April 2000 BRD regarding the relative performance of alternative methods would be helpful.

Appendix M of the BRD (page M-5, item 2, second sentence; formerly EPA Document 7 in **Appendix C**) appears to make an incorrect statement regarding the stopping rule. This Appendix discusses the stopping rule and suggests that "n," the number of animals, is always odd. The number of animals tested can be even (i.e., four) and may occur in three of the 11 possible testing sequences. The expression (n+1)/2 is equal to 2.5 for those sequences with four animals tested. Therefore, statements involving the expression (n+1)/2 are not always correct and require clarification.

The dosing section of the U.S. EPA Revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**) requires clarification regarding the actual procedure to be followed. The currently proposed procedure, described in the revised Guideline Section 23, line 5, states "if [the first] animal survives, two more animals are dosed sequentially at the limit dose." Since the Guideline requires that two more animals be tested regardless of outcome, the word "sequentially" should be deleted. Also regarding the revised Guideline, paragraph 23, line 6 states "if one or both of these two animals die, two animals are dosed sequentially at the limit dose…." However, conditions for stopping the test may be met after only one additional animal is tested. Therefore, the sentence should read, "if one or both of these two animals die, <u>additional</u> animals are dosed sequentially at the limit dose…." These two changes would help clarify the revised Guideline. This confusion can also be found in Appendix II, Paragraph 12 of the April 2000 Guideline, where the statement "then dose an additional two animals" is made; this statement is not always true and should be corrected. This type of statement is also mentioned in the April 2000 BRD (Section 2, 2.1.4, first paragraph). In the description of the testing scenarios in the April 2000 Guideline Appendix II, Paragraph 13, the

sequence S DD DX (in the most recent revision, O X XXU) is duplicated. There are only four sequences for this test that can end in death. Also, the parenthetical expressions can be eliminated because U would not occur in these sequences. All five of these sequences end with an S (or O in the most recent revision). Finally, in the April 2000 BRD (**Appendix C**, Tab 7, page C-184, first paragraph, third sentence), it is stated that the animals could be dosed sequentially or all at one time. The revised Guideline calls for dosing the animals sequentially--one at a time. This statement should be corrected. Consequently, the April 2000 Guideline and BRD provide a confusing and possibly contradictory description of dosing and should be corrected.

Due to the lack of clarity in the U.S. EPA Revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**), there appears to be a difference between the revised UDP Primary Test and the revised UDP Limit Test in the time of observation after dosing. The revised UDP Primary Test requires that the LD50 calculation be based on all reported deaths up to 14 days after dosing. The revised UDP Limit Test Guideline implies that decisions are based on all reported deaths that occur within two days. This discordance should be clarified by discussing the observation procedure as a general procedure in the revised Guideline. Currently, the observation period is only discussed in the paragraphs describing the revised UDP Primary Test.

While some features of the revised UDP Limit Test set it apart from the revised UDP Primary Test, most of the procedural steps for the two tests are similar. Consideration should be given to reorganizing the revised Guideline to improve clarity in a manner that indicates what features of the Guideline apply to both tests (e.g., test material preparation, dosing procedure, observation period, the intended range of materials amenable to the test, and testing of biopesticides). The April 2000 Guideline (Paragraph 17, page C-18) and the April 2000 BRD (Section 2.1.2.1, second sentence) do not provide adequate information regarding consideration of other acute toxicity data prior to conducting the test. However, this deficiency is common to all acute toxicity tests. Factors that pertain only to the revised UDP Limit Test should be clearly demarcated in the Limit Dose section of the revised Guideline. The Guideline should also state how to determine that a Limit Test and not the Primary Test is required.

4.3 Performance of the Revised UDP Limit Test

Information in the April 2000 BRD (such as in Sections 6.1, 6.3, and 6.5) was not helpful in determining if the revised UDP Limit Test adequately predicts whether the LD50 is above or below the limit dose. The only information identified for this task in the BRD was found formerly in EPA Document 7 in **Appendix C**, currently **Appendix M**. The performance of the revised UDP Limit Test was not tested with *in vivo* data, only with probability calculations. Based on the calculations, the procedure seems to work well and the performance characteristics may be adequate. However, it is not readily apparent how the revised UDP Limit Test was derived from these analyses. It would be helpful if the calculations were performed in a manner that allowed a clear comparison of the revised UDP Limit Test to the Limit Test described in OECD TG 401 (formerly **Appendix A**, currently **Appendix I**); instead, the calculations address the general issue of fixed versus sequential dosing.

The probability study (formerly in EPA Document 7 in **Appendix C**, currently **Appendix M**) begins with certain assumptions to be used for calculations. For example, the evaluation assumed that for all the animals tested there is the existence of a definable probit dose-response curve with a known LD50. However, if substantial variability exists in the animals during the study (e.g., in weight and age changes), there may not be a definable single slope. Weil et al. (1966) states that one of the more significant causes of laboratory-to-laboratory variability in estimates of the LD50 is the weight of the animals used. Because the April 2000 revised UDP Limit Test is a sequential procedure, the first animal tested will be younger and smaller than the last animal tested. There are no specific criteria given as to how wide the time span from the first to last animal tested can be for the test to remain valid. The primary concern is that the calculations

utilize a constant probability of death for a given level of exposure regardless of when that exposure occurs. This assumption is probably unrealistic given the sequential nature of the test and real life environmental factors that occur and can alter the probability of response during the conduct of the study.

With regard to the U.S. EPA Revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**), the Panel has several concerns regarding the accuracy of the revised UDP Limit Test and the ability of the test to minimize the use of animals. As indicated in the former **Appendix C**, Document 7, Table 3, current Table 3 in **Appendix M** on Page M-9, the probability of misclassification of a 5000 mg/kg UDP Limit Test for a sigma of 0.5 is 2% if the true LD50 is 1500. If the slope is more shallow, for example with a sigma of 2, the probability of misclassification of a 5000 mg/kg UDP Limit Test is increased such that a 21% misclassification occurs if the true LD50 is above 3000 mg/kg. Thus, there is concern about the accuracy of the revised UDP Limit Test, particularly for materials with shallow slopes for mortality. The table should be recalculated to provide the estimates for doses that represent the general Hazard Classes (i.e., 5 mg/kg, 50 mg/kg, 300 mg/kg, 2000 mg/kg, and 5000 mg/kg). This table would allow the reader to understand the chance of misclassifying various classes of toxic materials as non-toxic. Furthermore, similar comparisons using OECD TG 401 (formerly **Appendix A**, currently **Appendix I**) would clarify the strength of both tests. Additionally, the calculation that results in doses above 5000 mg/kg merits clarification in the April 2000 BRD.

The value of the revised UDP Limit Test would be improved if additional calculations were conducted regarding the probability for correct classification using other decision criteria. For example, assume failure of the revised UDP Limit Test when 1) any animal death occurs out of up to three tested, or 2) death of the first animal or death of two of five animals. These criteria may also yield a reduction in the number of animals tested. Consequently, additional calculations, similar to those in the revised BRD Table 3 in **Appendix M** on Page M-9, should be completed to determine if the expected number of animals tested is reduced.

The question of the need for additional calculations is discussed above. The April 2000 documentation did not provide *in vivo* studies to characterize the performance of the revised UDP Limit Test. It is laudable that probability calculations were used in an effort to help design a test procedure that would use fewer animals. However, it is not clear if the revised UDP Limit Test can be accepted in the absence of *in vivo* studies. Possibly, studies designed to test the practicability of the procedure, as was suggested for the revised UDP Primary Test, are needed.

The range of toxicity of the chemicals/products used to estimate the performance of the revised UDP Limit Test should be extended. The results from existing animal tests suggest it would probably help to have additional calculations using shallower slopes. It might be helpful to add results that would occur for LD50 values of 10000 and 20000 mg/kg. The additional information should provide a clearer picture of what occurs when materials with a fairly high LD50 are tested using this protocol. It would seem that materials with high LD50 values are those that would most likely be tested with the revised UDP Limit Test.

The April 2000 BRD (Section 2.5) describes the adequacy of results based on the explanation that a single experiment has been considered sufficient in the past. In general, this reasoning is not a scientifically sound justification for using only a single UDP Limit Test. The adequacy of a single experiment is not a major factor that needs to be considered since the purpose of the UDP Limit Test is to provide the same information as past testing while reducing animal use.

4.4 Reliability (Intra-laboratory Repeat-ability, Inter-laboratory Reproducibility) of the Revised UDP Limit Test

In vivo acute lethality data were not considered in the evaluation of the reliability of the revised UDP Limit Test. The only available data are based on probability calculations shown in the revised BRD Table 3 in **Appendix M**, Page M-9 of the BRD.

The problems associated with this approach are discussed above.

With regard to the revised UDP Guideline, the only scientific basis for the revised UDP Limit Test is the probability calculations. Much of the April 2000 BRD documentation does not appear to apply to the revised UDP Limit Test. Extrapolating from studies used to estimate the LD50, it appears that the revised Guideline must be specific in all aspects of study design in order to ensure adequate LD50 reproducibility. The Guideline may not be sufficiently specific to ensure reproducibility. Factors such as the age and weight of the animals used appear to be very important to ensuring adequate reproducibility, but these factors are not rigorously specified in the revised Guideline. The specific determination of whether an animal is moribund and should be humanely killed can vary from investigator to investigator. Because no more than five animals will be used, an error in a single observation can have a major influence on outcome. Only *in vivo* studies appear able to address these issues.

4.5 Summary Conclusions

With regard to the U.S. EPA Revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**), the Panel members reviewing the revised UDP Limit Test concluded that the test has been evaluated sufficiently. Its performance is satisfactory to support its adoption as a substitute for the Limit Test described in OECD TG 401 (formerly **Appendix A**, currently **Appendix I**) for oral acute toxicity. However, there are qualifications regarding the accuracy and reliability of the Limit Test. The revised UDP Limit Test is expected to perform as well as or better than the Limit Test in OECD TG 401, with a reduction in the number of animals. Regarding animal welfare, the Panel members also discussed whether the revised UDP Limit Test adequately considered and incorporated where scientifically feasible, procedures that refine, reduce, and/or replace animal use. The revised UDP Limit Test does not replace animal use. It was not clear to these Panel members if the procedure refined animal use, in terms of reducing pain and suffering. However, the majority of these Panel members concluded that the procedure reduced animal usage, particularly in comparison to the Limit Test in OECD TG 401.

The Panel members noted deficiencies in the description of the revised UDP Limit Test in the April 2000 Guideline and BRD. The scientific basis for the revised UDP Limit Test is not adequately described in either document. There was no rationale provided for the method. Little justification for the UDP Limit Test is provided in the BRD, particularly regarding the starting dose (i.e., 2000 or 5000 mg/kg). The overall product was inadequately organized for review of the revised UDP Limit Test. The revised UDP Limit Test Guideline was not well written and the organization of the current document made it difficult to locate the relevant sections to address the questions in the Evaluation Guidance. The relationship of the revised UDP Limit Test to the revised UDP Primary Test is unclear in the April 2000 BRD. The probability calculations and presented data were insufficient to determine the accuracy for correct classification at shallow slopes. Other limitations of the revised UDP Limit Test are also present in the revised UDP Primary Test and in acute toxicity testing in general.

4.6 Recommendations

1. The scientific basis of the revised UDP Limit Test should be included in the U.S. EPA Revised UDP Guideline (formerly **Appendix C**, currently **Appendix G**), with greater explanation in the April 2000 BRD.
2. Additional discussion in the revised Guideline of the applicability of the UDP Limit Test in hazard or safety assessment would significantly strengthen the test. A decision criteria flow chart describing the complete testing scheme might be an efficient way to achieve this goal.
3. The revised Guideline would be improved if a short rationale for the UDP Limit Test were added in a separate paragraph.
4. The revised Guideline as currently written is difficult to follow. Consideration should be given to reorganizing the Guideline to improve clarity.

5. The use of constant volume or constant concentration of the test material should be allowed.
6. In the Guideline, all reference to littermates should be excluded.
7. Animals of 8 to 12 weeks of age at the time of dosing should be used.
8. The individual animal body weights on the day of dosing must be within 20% of the mean body weight for all animals dosed.
9. Clarification of the selection of the limit dose would be helpful in the April 2000 Guideline and BRD.
10. The current organization of the BRD made adequate document evaluation difficult. Movement of some material in former **Appendix C**, Tab 7 (current **Appendix M**) to the main section of the BRD would improve the organization and address many issues of concern. Furthermore, clarification of several details in the Guideline or the BRD would improve the understanding of the test.
11. Additional calculations to justify the benefits of the revised UDP Limit Test would be helpful. The document should provide probability estimates for accuracy using criteria that compare the revised UDP Limit Test to OECD TG 401 (formerly **Appendix A**, currently **Appendix I**) to clearly delineate the benefits. The document should provide probability estimates for accuracy using more stringent criteria to determine if a further reduction in the number of animals tested is possible.
12. Table 3 in former **Appendix C**, Document 7 (current **Appendix M** on Page M-9) should be recalculated to provide dose estimates that represent the general Hazard Classes (i.e., 5 mg/kg, 50 mg/kg, 300 mg/kg, 2000 mg/kg, and 5000 mg/kg). It might be helpful to add results that would occur for LD50 values of 10000 and 20000 mg/kg.
13. The value of the revised UDP Limit Test would be improved if additional calculations were conducted regarding the probability for correct classification using other decision criteria.
14. The basic stopping rule in the revised UDP Limit Test is the occurrence of two additional survivors or three deaths following survival of the first animal. This rule differs from the stopping rule applied when reaching the upper bounding limit during the revised UDP Primary Test, which requires that three consecutive animals must survive. The two different stopping rules may cause confusion and additional explanation in the BRD is suggested to address this issue.

5.0 UDP SUPPLEMENTAL TEST TO ESTIMATE SLOPE AND CONFIDENCE INTERVALS

5.1 Introduction and Rationale for the UDP Supplemental Test

While there are several reasons why some estimate of the slope for the dose-response curve may be needed, none were articulated in the BRD. Slope information is, for example, useful in selecting doses for subsequent longer-term studies. However, determination of an exact slope is rarely necessary.

One exception is that the U.S. EPA has a legal requirement to perform wildlife risk assessments for acute toxicity. Within the 29 countries of the OECD, this exception appears to be the only regulatory requirement for a rodent acute toxicity test that generates the slope of the dose-response curve as well as an LD50 value. It is uncertain what proportion of all acute toxicity tests will be required by the U.S. EPA to provide a slope value. Will it only apply to new pesticide active ingredients or will such information also be needed for all new formulations being registered for use? Is the inclusion of the UDP Supplemental Test in the revised OECD TG 425 justified? Far fewer animals would be killed if information on slope were requested through the conduct of a non-guideline study. A non-guideline study could utilize any scientifically relevant test method, as agreed upon by the registrant and the Agency. The revised OECD TG 425 would then contain only the acceptable UDP Primary and Limit Tests and would allow the OECD to proceed with the deletion of OECD TG 401 (formerly **Appendix A**, currently **Appendix I**) and approval of a method that further reduces animal use for acute toxicity testing.

The scientific basis for the proposed UDP Supplemental Test is not adequately described or even addressed. Why and when such data would be needed is not defined. The justification for the UDP Supplemental Test presented in the BRD is discussed in statistical terms stating that the UDP proposed by Dixon and Moods (1948) centers trials around the LD50 value. This method is appropriate for estimating the LD50, but it is not a good means of estimating the 'slope' in the probit model. The fit of the UDP Supplemental Test into a strategy for hazard or safety assessment is not adequately discussed. The lack of a description of the utility of this test in hazard assessment was a significant omission.

The BRD makes the point that more animals are needed for the generation of sound data for determining slope and confidence intervals (CI) for LD50s. This requirement is a fundamental problem with the proposed UDP Supplemental Test—too few data points. This issue makes it very questionable that the proposed UDP Supplemental Test would meet published regulatory acceptance criterion that "the method should be suitable for international acceptance." To increase the number of animals used per test, without demonstrated and necessary improvements in precision, would not be consistent with the regulatory acceptance criterion that "the method must provide adequate consideration for the reduction, refinement, and replacement of animal use." Compared to OECD TG 401 (formerly **Appendix A**, currently **Appendix I**), the proposed UDP Supplemental Test meets the criterion for reduction in that it provides better quality information from fewer animals.

Virtually no information was provided that would allow a determination on whether the intended range of materials, based on chemical class or physico-chemical factors, was appropriate. As noted in the Summary Conclusions, the number of agents tested, the number of chemical classes evaluated, and the range of effects expected are far fewer than what would be needed to adequately address this question. Additional background information is needed to properly evaluate any new procedure proposed to generate slope and CI information in addition to the LD50 value.

The slope is said to be equal to 1/sigma (in one place the BRD says proportional to 1/sigma), but is never directly defined. What is 1/sigma the slope of? The definition of slope should be clearly provided in the Guideline and in the BRD upon the first mention of slope. The slope of a

probit curve is a different value at each point on the curve.

What scientific questions are being asked where the "slope" is required for determining the answers? Information of this type in the BRD is too vague. For example, in U.S. EPA Document 1, page 9, it states that, "Some authorities also use test results to perform various risk assessment functions, including determination of confidence interval and slope to make projections at the low end of the dose-response curve." The Panel was unable to discern what data need would be satisfied by the calculation of slope and CI, or how low on the dose-response curve that data points would be extracted.

If the slope is being used to estimate the LDp, where p is some toxicity rate other than 50%, then what values of p are being used and for what purposes? The BRD presents one example in which 20% of the LD50 is of interest. This example is odd in that the toxicity rate associated with 0.2 LD50 depends on the steepness of the probit curve and has no intrinsic meaning. Furthermore, there is a problem with the regulations and/or procedures that use criteria based on k*LD50, such as are reported in Federal Regulation (40 CFR(129)). It needs to be emphasized that k*LD50 is not LD(k*50). For example, 1/10*LD50 is not the dose at which the chemical is toxic for 1/10*50=5 percent of the population. The basis for this convention of setting standards at k*LD50 is incomprehensible because the toxicity rate at this level depends entirely on the slope of the dose-response curve and does not provide a constant standard in obvious manner. Criteria for toxicity should be stated in terms of the LDp, where p is between 0 and 1, and presumably less than or equal to 0.5.

The level of precision required for the estimates of slope and CI should be stated. This information is important because a procedure that is efficient for one objective is likely to be less efficient for a different objective. A toolbox of procedures is needed to meet different objectives. For example, a good procedure for estimating the LD50 and the slope will not be so helpful in estimating the LDp for p far from 50. The latter would require the correct model and extremely good precision. The consequences of using a procedure for anything but its designed purpose need to be presented. The BRD should clarify whether a CI is for the LD50, the slope, or if both are needed. It should also be stated how the CI is to be calculated and interpreted.

Although not explicitly stated, it appeared to the Panel that there was a lack of distinction between the CI for the LD50 and certain percentiles of the probit curve. These two need to be clearly defined in the Guideline to avoid confusion. In particular, if exposures were selected independently and randomly from a normal density, a 95% CI for the LD50 would be the estimated LD50 +/- 1.96*sigma/sqrt(n), where n is the sample size. However, in none of the procedures (1987 OECD TG 401, OECD TG 425, or the revised UDP; **Appendices I, H,** or **G**, respectively) are exposures selected randomly from a normal density. Thus, the use of the constant 1.96 in establishing a CI for the LD50 is arbitrary and not related in any know manner to some degree of confidence. In fact, the LD50 +/- 1.96 sigma gives estimates of the LD2.5 and the LD97.5. The CI for the LD50 using the UDP and its revision will depend on the interval between doses as well as on sigma. The formula for the CI of the LD50 also will depend on the type of estimator (e.g., Maximum Likelihood Estimate (MLE) or Modified Isotonic Estimate (MIE)) and the procedural rules that prescribe how exposures are selected.

The CI for the LD50 given maximum likelihood estimation can be obtained using an expression for the variance of the estimated LD50 that is given, for example, by Mats et al. (1998). It could also be obtained from replicated experiments or bootstrapping [See Stylianou (2000), for details on bootstrapping the CI of the LD50].

From the simulations, the dose progression proposal appears to be efficient for estimating the slope when it is high, but not when the slope is low. Furthermore, few animals are tested at doses far from the LD50, therefore, the efficiency level for this procedure is not maximized. In the BRD (U.S. EPA Document 8, Part D), it is shown that treating near, but not at, the optimal dose can result in significantly reduced efficiency. A slight

modification of the UDP as described in the April 2000 Guideline Appendix II (formerly **Appendix C**, currently **Appendix G**) will cluster the exposures around the optimal doses, even though they are unknown. We anticipate that other starting and stopping rules, as well as a dose progression schedule, can be developed to improve the current proposed UDP Supplemental Test, as well as the 1987 OECD TG 401 (formerly **Appendix A**, currently **Appendix I**) procedure.

5.2 UDP Supplemental Protocol

The general description is unclear as written. The complicated, statistically-based language is difficult to comprehend and translate into a manageable protocol, even by an experienced study director. More detail is needed and an example of the procedure (i.e., showing dose levels with response/no response) would be helpful. Potential problems exist where the Guideline makes statements such as "based on results, good judgement is required" and a possible "alternate procedure" may be appropriate. Also, an explanation for “staggered” starting doses is needed. The use of other acute toxicity information is mentioned, but is neither discussed nor is its relevance to dose setting addressed.

Computer simulations were used to consider possible outcomes of the UDP Supplemental Test and these simulations seem adequate. However, this approach is no substitute for actual laboratory studies. Comments from laboratory personnel who conduct these studies routinely should be carefully considered. Not only should the predictability of the test be considered, but also the difficulty involved in conducting the test. This procedure would require constant monitoring of responses and identification of each next dose, followed by a relatively complicated computer analysis for slope and CI.

The UDP Supplemental Test will take longer to complete as compared with a standard LD50 OECD TG 401 study (formerly **Appendix A**, currently **Appendix I**). A time of 48 hours between each dosing must be used. If dosing was performed on Monday, Wednesday, and Friday (requiring observations on Saturday and Sunday), and 15 animals were needed, the test would take at least five weeks to complete. The UDP Supplemental Test would require at least another five weeks, for a total of at least 10 weeks. This is a relatively long time period for conducting an acute oral toxicity study. Industry is attempting to shorten development timelines for new chemicals as much as possible and an additional month of testing for an acute oral LD50 study could be significant. In addition, the need to test large numbers of chemicals, as in the High Production Volume chemicals program, will result in testing laboratories quickly reaching capacity. The time to complete these studies should be considered.

There are major concerns over the practicality of performing the UDP Supplemental Test in a standard toxicology laboratory. To ensure that the age/weight range is not exceeded late in the testing period, the number of animals required at study initiation could be quite high. Many of these could be wasted if other tests were not being conducted in the laboratory over the same period. Hence, not only does the UDP Supplemental Test procedure use no fewer animals than the OECD TG 401 procedure, it could indirectly result in the death of more animals because unused animals may have to be culled.

While, on the surface, the UDP Supplemental Test appears quite simple to conduct, the uncertainties that may be involved make it far from simple. Moreover, because the UDP Supplemental Test has never actually been conducted *in vivo*, the question of whether the general procedures are appropriate and described in sufficient detail cannot be ascertained.

5.3 Performance of the UDP Supplemental Test based on Computer Simulations

Based only on computer simulations, the usefulness of the UDP Supplemental Test cannot be determined without better knowledge of its intended purpose. The numbers and types of chemicals represented by the simulations were not appropriate. Reference was made to a listing of data from six pesticides, but there was no indication in the BRD as to where this information was used. The range of dose-response curves presented seemed adequate; however, very

shallow or steep dose-response curves should have been discussed in greater depth.

There was little evidence that the developers attempted to summarize the results from the large number of simulations. The description of Simulations II and III of BRD former U.S. EPA Document 8, Part D (current **Appendix N-4**), states that "for each run the computer randomly picked the appropriate number of animals from the entire population …". What is this population? Is it assumed that the animals are normally distributed around the LD50, with standard deviation sigma, and if so, why would this be the case? A population of very sensitive animals might be concentrated around the LD85, for example. If some other distributional assumptions were made, what are they?

5.4 Reliability (Intra-Laboratory Repeat-ability, Inter-Laboratory Reproducibility) of the UDP Supplemental Test

A major weakness of the proposed UDP Supplemental Test is that no confirmatory testing against conventional *in vivo* studies has been conducted. Any conclusions regarding the reliability of the UDP Supplemental Test are significantly restricted by the absence of *in vivo* data. The premise that computer simulations alone are sufficient for predicting biological events is not accepted by most scientists in the life sciences arena.

The issue of intra- and inter-laboratory variability has not been adequately addressed for the UDP Supplemental Test protocol. This failure is a major reason for a lack of confidence in this procedure. Some inter-laboratory variability is inherent in any test and information in the BRD indicates that values obtained with the standard LD50 study can vary by at least three-fold. There have been no inter-laboratory variability comparisons for the revised UDP Primary Test or for the UDP Supplemental Test. With the UDP Supplemental Test, additional variability may result from the fact that the rats tested may be of different weights/ages due to the length of testing. Also, the timeline for waiting for animal deaths to occur may add variability. Some investigators may dose animals every 48 hours to accelerate the process, while others may wait longer between dosing to better assess for delayed deaths.

5.5 Summary Conclusions

1. The UDP Supplemental Test for slope and CI was not recommended for adoption. The Panel was unable to evaluate the utility of the test because sufficient information regarding the use of the data was not provided.
2. The revised UDP Primary Test and Limit Test adequately consider and incorporate procedures that reduce animal use. For the revised UDP Primary Test, the use of 0.5 log units for dose spacing is reasonable and appropriate based on experience and the results of computer modeling. This spacing allows the investigator to move through dose levels more quickly and thereby limits the number of animals used. In contrast, the UDP Supplemental Test, which includes the determination of slope, may use more animals than OECD TG 401 (formerly **Appendix A**, currently **Appendix I**). The UDP Supplemental Test does not replace animal use. Because the UDP Supplemental Test requires the use of starting doses below the LD50, there is a possibility that overall pain and distress may be reduced compared to OECD TG 401. At this point, there are no alternative animal species more suitable than rats for obtaining the type of information generated in acute toxicity testing.
3. The development of the UDP Supplemental Test has not followed the customary track for evaluating alternative methods in that only computer simulations were conducted. No actual *in vivo* testing was performed.
4. It is acknowledged that there has been a desire for a number of years to delete OECD TG 401, primarily for humane reasons. It is clear that the revised UDP Primary Test is an attractive replacement along with the revised UDP Limit Test, the FDP, and the ATC methods for estimating acute toxicity. While the UDP Supplemental Test was designed and proposed as a means of estimating the slope and CI, it is not clear whether this design is appropriate to address regulatory data needs. Moreover, these data needs have not been clearly presented to the Panel.

5. The BRD would be improved by closer attention to the norms of good method development and a clearer, more focused document preparation.
6. In Guideline Section 13.0 (UDP Supplemental Protocol) and in Addendum III of the Panel Report (Statistical Evaluation of the Revised UDP and the UDP Limit Test), a number of suggestions are offered that may be evaluated by the sponsors of this peer review.
7. If a procedure is needed to define points on the dose-response curve well below the median lethal dose, an alternative procedure, such as that detailed in Addendum I of this Report (Direct Estimation of a Point on the Dose-Response Curve that is far from the LD50), can be considered. Similarly, one possible alternative method for calculating the slope is presented in Addendum II of this Report (Consideration for Estimating the Slope).

5.6 Recommendations

1. Regulatory data needs currently addressed by estimation of the slope and CI derived from acute oral toxicity studies in the rat and other species need to be more clearly defined.
2. Consideration should be given as to whether the slope and CI are the most appropriate parameters for addressing regulatory data needs or if these needs can be addressed more directly. For example, an alternative procedure outlined in Addendum I of this Report may be used to estimate points on the dose-response curve well below the median lethal dose.

6.0 REFERENCES

16 CFR 1500. 2000. Title 16: Commercial Practices. Chapter II. Consumer Product Safety Commission. Part 1500: Hazardous Substances and Articles; Administration and Enforcement Regulations. Government Printing Office, Washington, DC.

29 CFR 1910.1200. 1998. Title 29: Department of Labor. Chapter XVII. Part 1910: Occupational Safety and Health Administration. Subpart Z: Toxic and Hazardous Substances. Section 1200: Hazard Communication. Government Printing Office, Washington, DC.

40 CFR 156. 2000. Title 40: Protection of Environment Agency. Code of Federal Regulations. Part 156: Labeling Requirements for Pesticides and Devices. Government Printing Office, Washington, DC.

49 CFR 173. 1999. Title 49: Department of Transportation. Code of Federal Regulations. Part 173: Shippers--General Requirements for Shipments and Packagings. Government Printing Office, Washington, DC.

American Society for Testing and Materials (ASTM). 1987. Standard Test Method for Estimating Acute Oral Toxicity in Rats. ASTM E1163-87. In: Annual Book of ASTM Standards, Philadelphia.

Barlow, R.E., D.J. Bartholomew, J.M. Brenner, and H.D. Brunk. 1972. Statistical Inference Under Order Restrictions: The theory and application of isotonic regression. John Wiley & Sons, New York. 388 pp.

Bonnyns, E., M.P. Delcour, and A. Vral. 1988. Up-and-Down Method as an Alternative to the EC-Method for Acute Toxicity Testing. IHE Project No. 2153/88/11. Institute of Hygiene and Epidemiology, Ministry of Public Health and the Environment, Brussels. 33 pp.

Bruce, R.D. 1987. A Confirmatory Study for the Up-and-Down Method for Acute Toxicity Testing. Fundam. Appl. Toxicol. 8:97-100.

Bruce, R.D. 1985. An Up-and-Down Procedure for Acute Toxicity Testing. Fundam. Appl. Toxicol. 5:151-157.

Dixon, W.J. 1991. Staircase Bioassay: The up-and-down method. Neurosci. Biobehav. Rev. 15:47-50.

Dixon, W.J. 1965. The Up-and-Down Method for Small Samples. J. Am. Stat. Assoc. 60:967-978.

Dixon, W.J., and A.M. Mood. 1948. A Method for Obtaining and Analyzing Sensitivity Data. J. Am. Stat. Assoc. 48:109-126.

Durham, S.D., and N. Flournoy. 1995. Up-and-Down Designs I: Stationary treatment distributions. In: Adaptive Designs, Flournoy, N. and W.F. Rosenberger (Eds.). Hayward, California: Institute of Mathematical Sciences. pp. 139-157.

Durham, S.D., and N. Flournoy. 1994. Random Walks for Quantile Estimation. In: Statistical Decision Theory and Related Topics V, Gupta, S.S., and J.O. Berger (Eds.). New York: Springer-Verlag. pp. 467-476.

Durham, S.D., N. Flournoy, and A.A. Montazer-Haghighi. 1995. Up-and-Down Designs II: Exact treatment moments. In: Adaptive Designs, Flournoy, N., and W.F. Rosenberger (Eds.). Hayward, California: Institute of Mathematical Sciences. pp. 158-178.

Durham, S.D., N. Flournoy, and W.F. Rosenberger. 1997. A Random Walk Rule for Phase I Clinical Trials. Biometrics 53:745-760.

Flournoy, N. 1993. A Clinical Experiment in Bone Marrow Transplantation: Estimating a percentage point of a quantal response curve**.** In: Case Studies in Bayesian Statistics, Gatsonis, C, J.S. Hodges, R.E. Kass, and N.D. Singpurwala (Eds.). New York: Springer-Verlag. pp.324-336.

Galson, S. 2000. Historical and current regulatory perspectives. Opening Plenary Session, ICCVAM International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity, October 17-20, 2000.

Griffith, J.F. 1964. Interlaboratory Variations in the Determination of Acute Oral LD50. Toxicol. Appl. Pharmacol. 6:726-730.

ICCVAM. 1997. Validation and Regulatory Acceptance of Toxicological Test Methods: A report of the *ad hoc* Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication 97-3981. National Institute of Environmental Health Sciences, Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/docs/guidelines/validate.pdf [cited October 18, 2001].

Levitt, H. 1971. Transformed Up-Down Methods in Psychoacoustics. J. Acoustical Soc. America 49:467-447.

Lipnick, R.L., J.A. Cotruvo, R.N. Hill, R.D. Bruce, K.A. Stitzel, A.P. Walker, I. Chu, M. Goddard, L. Segal, J.A. Springer, and R.C. Myers. 1995. Comparison of the Up-and-Down, Conventional LD50, and Fixed-Dose Acute Toxicity Procedures. Food Chem. Toxicol. 33:223-231.

Mats, V.A., W.F. Rosenberger, and N. Flournoy. 1998. Restricted Optimality for Phase I Clinical Trials. In: New Developments and Applications in Experimental Designs, Flournoy, N., W.F. Rosenberger, and W.K. Wong (Eds.). IMS Monograph Series 34:50-61.

Mulder, G.J. 1986. Sex Differences in Drug Conjugation and their Consequences for Drug Toxicity. Sulfation, glucuronidation and glutathione conjugation. Chem. Biol. Interactions 57:1-15.

National Institute of Environmental Health Sciences (NIEHS). 2000a. National Toxicology Program: Request for Data and Nomination of Expert Scientists to Participate in the Independent Peer Review Evaluation of the Revised Up-and-Down Procedure for Assessing Acute Oral Toxicity. Evaluation of the Up-and-Down Procedure. 65 FR 8385. February 18, 2000.

NIEHS. 2000b. National Toxicology Program: Notice of Peer Review Meeting on the Revised Up-and-Down Procedure (UDP) as an Alternative Test Method for Assessing Acute Oral Toxicity. Request for Comments. 65 FR 35109. June 1, 2000.

Nelson, D.R., L. Koymans, T. Kamatski, J.J. Stegeman, R. Feyereisen, D.J. Waxman, M.R. Waterman, O. Gotoh, M.J. Coon, R.W. Estrabrook, I.C. Gunsalus, and D.W. Nebert. 1996. P450 Super Family: Update on new sequences, gene mapping accession numbers and nomenclature. Pharmacogenetics 6:1-42.

Organisation for Economic Co-operation and Development (OECD). 2001. Harmonised Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixture. OECD Chemicals Committee and the Working Party on Chemicals, Pesticides, and Biotechnology, Series on Testing and Assessment, No. 33. OECD, Paris. 247 pp. Available: http://www.oecd.org/ehs/class/HCL6.htm. [cited October 18, 2001].

OECD. 2000. Guidance Document on the Recognition, Assessment, and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation, OECD Environmental Health and Safety Publications, Series on Testing and Assessment, No. 19. OECD, Paris. 44 pp. Available: http://www.oecd.org//ehs/test/monos.htm. [cited October 18, 2001].

OECD. 1999a. OECD Guideline for Testing Chemicals Revised 420: Acute Oral Toxicity - Fixed Dose Procedure. OECD, Paris.

OECD. 1999b. OECD Guideline for Testing Chemicals Revised 423: Acute Oral Toxicity-Acute Toxic Class Method. OECD, Paris.

OECD. 1998a. OECD Guideline for Testing Chemicals 425: Acute Oral Toxicity: Up-and- Down Procedure. OECD, Paris.

OECD. 1998b. Harmonized Integrated Hazard Classification System for Human Health and Environmental Effects of Chemical Substances as Endorsed by the 28th Joint Meeting of the Chemicals Committee and the Working Party on Chemicals in November 1998, Part 2, p. 11. Available: http://www.oecd.org/ehs/class/HCL6.htm. [updated August 14, 2001 -- see also OECD, 2001].

OECD. 1987. OECD Guideline for Testing Chemicals Test Guideline 401: Acute Oral Toxicity. OECD, Paris.

Schlede, E., W. Diener, U. Mischke, and D. Kayser. 1994. Organisation for Economic Co-operation and Development expert meeting: Acute toxic class method. January 26-28, 1994, Berlin, Germany.

Schlede, E., U. Mischke, W. Diener, and D. Kayser. 1995. The International Validation Study of the Acute Toxic Class Method (oral). Arch. Toxicol. 69:659-670.

Schlede, E., U. Mischke, R. Roll, and D. Kayser. 1992. A National Validation Study of the Acute Toxic Class Method - An alternative to the LD50 test. Arch. Toxicol. 66:455-470.

Sitter, R.R. and C.F.J. Wu. 1993. Optimal Designs for Binary Response Experiments: Fieller, D and A criteria. Scandinavian J. Statistics 20:329-341.

Robertson, T., F.T. Wright, and R.L. Dykstra. 1988. Order Restricted Statistical Inference, John Wiley & Sons, New York.

Spielmann, H., E. Genschow, M. Liebsch, and W. Halle. 1999. Determination of the Starting Dose for Acute Oral Toxicity (LD50) Testing in the Up-and-Down Procedure (UDP) from Cytotoxicity Data. ATLA 27:957-966.

Stylianou, M. 2000. A New Approach to Dose Finding for Phase I Clinical Trials. Dissertation. American University.

Stylianou, M., and N. Flournoy. 2000. A New Approach to Dose Finding for Phase I Clinical Trials. Technical Report Number 2000-2. Department of Mathematics and Statistics. American University.

Trevan, J.W. 1927. The Error of Determination of Toxicity. Proc. Royal Soc. 101B:483-514.

van den Heuvel, M.J., D.G. Clark, R.J. Fielder, P.P. Koundakjian, G.J.A. Oliver, D. Pelling, N.J. Tomlinson, and A.P. Walker. 1990. The International Validation of a Fixed-Dose Procedure as an Alternative to the Classical LD50 Test. Food Chem. Toxicol. 28:469-482.

van den Heuvel, M.J., A.D. Dayan, and R.O. Shillaker. 1987. Evaluation of the BTS Approach to the Testing of Substances and Preparations for their Acute Toxicity. Human Toxicol. 6:279- 291.

Weil, C.S. 1983. Economical LD50 and Slope Determinations. Drug Chem. Toxicol. 6:595-603.

Weil, C.S. 1975. Toxicology Experimental Design and Conduct as Measured by Interlaboratory Collaborative Studies. J. Off. Anal. Chem. 58:683-688.

Weil, C.S., C.P. Carpenter, and H.F. Smyth. 1953. The Median Effective Dose. Ind. Hyg. Q. 14:200-206.

Weil, C.S., C.P. Carpenter, J.S. West, and H.F. Smyth. 1966. Reproducibility of Single Oral Dose Toxicity Testing. Am. Ind. Hyg. Assoc. J. 27:483-487.

Weil, C.S., and G.J. Wright. 1967. Intra- and Inter-laboratory Comparative Evaluation of a Single Oral Test. Toxicol. Appl. Pharm. 11:378-388.

Yam, J., P.J. Reer, and R.D. Bruce. 1991. Comparison of the Up-and-Down Method and the Fixed Dose Procedure for Acute Oral Toxicity Testing. Food Chem. Toxicol. 29:259-263.

Zbinden, G., and M. Flury-Roversi. 1981. Significance of the LD50 Test for the Toxicological Evaluation of Chemical Substances. Arch Toxicol. 47:77-99.

Addendum I: Direct Estimation of a Point on the Dose-Response Curve That Is Far From the LD50

Estimating a LDp value that is near the LD50 is quite robust with respect to model assumptions; however, sensitivity increases as the LDp of interest moves away from the LD50. This increase in sensitivity is as expected because typical models (e.g., logistic, probit, Weibull) differ most in the tails. Relying on estimates of model parameters to estimate a low (high) LDp with only a few animals should and can be avoided by using a nonparametric procedure with a nonparametric estimator.

Exposures can be tailored to cluster around an unknown LDp, such as the LD16, using a slight modification of the UDP called the Biased Coin Up-and-Down Design (BCD) [Durham and Flournoy, 1994; see also Durham et al., 1997].

By using the BCD with any increasing dose-response function, such as the probit, exposures will quickly cluster around any target LDp, similar to what the standard UDP does for the LD50. To cluster points around the LD1p, $p \leq 0.50$, proceed as follows:

> Use a biased coin, with probability of heads $=[p/(1-p)]$. If there is a toxic response, treat the next animal at the next lower dose; if there is a non-toxic response, flip the biased coin. If the coin comes up tails, treat the next sequential animal at the same dose; if the coin comes up heads, treat the next sequential animal at the next higher dose.

Note that for $p=0.50$, the BCD procedure reduces to Dixon and Mood's (1948) up-and-down design. For $p>0.50$, see Durham and Flournoy (1994).

The Modified Isotonic Estimate (MIE) of the LDp, described in Addendum IV, is an attractive alternative to the Maximum Likelihood Estimate (MLE) since it does not require a probit or other parametric model assumption. This approach is particularly important for estimating a LDp far from the LD50 where model differences are most pronounced. Stylianou and Flournoy (2000) demonstrate that the MIE outperforms other nonparametric estimators found in the literature, and compares well with the MLE.

It appears that no one asked how accurately OECD TG 401 (formerly **Appendix A**, currently **Appendix I**) provided estimates of toxicity at low doses, using the estimation of the slope in a probit model; however, the Panel was asked to evaluate the UDP Supplemental Test protocol for estimating toxicity rates at fractions of the LD50. Finding that little thought had been given to precision, our evaluation cannot determine whether this requirement will be met. Some consideration should be given to stopping rules that take precision into account. Stylianou (2000) considered stopping rules for the BCD. A likelihood ratio test similar to Rule #3 in the revised UDP Primary Test may work well also. This approach should be evaluated.

Addendum II: Considerations for Estimating the Slope

The "optimal design" (i.e., the procedure yielding the most information about the LD50 and the slope simultaneously, with a fixed number of animals) would be to administer the test substance to animals (*cf.* Sitter and Wu, 1993) at the:

- LD13 and LD87 if the response function is probit,
- LD18 and LD82 if the response function is logistic,
- LD10, 50, and 90 if it is double exponential, and
- LD21, 50, and 79 if it is double reciprocal.

A compromise might be to treat animals at LD16, LD50, and LD84 (if possible). If avoiding highly toxic doses is desired, the LD16 and LD50 are attractive choices. Assuming a probit dose-response function, the LD16 and LD84 are –1 and +1 sigma from the LD50, respectively. Thus, the estimates of sigma can be obtained from estimates of [LD84-LD16]/2, [LD84-LD50], and [LD50-LD16]. Differences in these estimates would indicate that the sample size is too small or that the probit model is not a good fit.

As recognized by the development team for the revised UDP, even assuming the probit model, it is impossible to implement the optimal design because the optimal values of LDp are unknown. Certainly, selecting a few dose levels (based on certain expectations as in OECD TG 401) and treating a fixed number of animals at those dose levels can be very inefficient, because even good expectations based on considerable experience can be incorrect (see, for example, Flournoy, 1993). Simulations in BRD U.S. EPA Document 8, Part D demonstrate also the decline in efficiency that can result from the use of designated points near, but not at, the optimal ones.

To deal with this efficiency issue, the UDP Supplemental Procedure incorporates several escalation-dosing series, starting at low doses. The problem with increasing the dose at every nontoxic outcome is that exposures are closer to the LD50 than to doses such as the LD16 after only a couple of animals.

Simulations in former U.S. EPA Document 8, Part D (current **Appendix N-4**) indicate that the UDP Supplemental Test procedure yields a reasonable estimate of sigma when sigma is small, but substantially underestimates sigma when sigma is large. This discrepancy could result from the dose escalation procedures when very few animals are tested at levels far from the LD50, or because of the large interval between doses. These two possibilities should be examined.

To shorten the time required for estimating the LD50 and slope together, simultaneously conducting BCD procedures to target two or three points on the dose-response curve (e.g., the LD16 and LD50, the LD16 and LD84, or the LD16, LD50, and LD84) should be considered. Clustering treatments around but not at two or three nearly optimal dose levels using simultaneous BCD is expected, on theoretical grounds, to produce more efficient estimates of the LD50 and slope when compared to the UDP Supplemental Test.

MIE (see Addendum IV of this report) of the necessary LDp values are attractive alternatives to MLE. Of course, more animals are required to estimate LDp values distant from the LD50, but at least for doses as low as the LD10, the expected increase in the number of animals is modest. In particular, the expected number of animals required is less than that required by the combined UDP Primary and Supplemental Tests for estimating both the LD50 and sigma. Additionally, targeting the LD16 and the LD50 will be less efficient for estimating sigma and the LD50 than targeting the LD16, LD50, and LD84, and also much less efficient than targeting only the LD16 and the LD84. The relative efficiency of targeting the three points versus two points on the dose-response curve should be examined. For example, it could take many more animals targeting two dose levels (instead of three) to get the same quality estimates of the LD50 and sigma. If animals should not be treated around the LD84 to avoid pain and suffering, this point is moot.

Addendum III: Summary of the Statistical Evaluation of the Revised UDP

Significantly more information per animal will be obtained using an up-and-down procedure for estimating the LD50 when compared to treating fixed numbers of animals at several doses. This increase in the extent of information per animal has been shown theoretically (*cf.* former references 1-6 of U.S. EPA Document 2, current **Appendix J-2**) and has been demonstrated in the simulation studies provided in the BRD. A suggestion to simplify the use of the likelihood ratio statistic as a stopping rule is offered for consideration by the development team.

It is important to recognize that the variability of the LD50 estimate increases with the step size used between sequential dose levels. The UDP is proposed for many different purposes and varying degrees of precision will be appropriate for different purposes. For example, for the crude classification of chemicals, a large dose progression factor with its associated relatively large variation in the LD50 estimate will be satisfactory. However, when considering the effect of a chemical on an endangered species, considerably greater precision is desired. One may predict that the precision expected for some purposes simply cannot be obtained with the proposed step size. To prepare for a revision (perhaps three years from now), it is recommended that the precision desired for different purposes be ascertained. This information would be used to develop rules for adjusting the step size (and perhaps the nominal sample size and stopping criteria as well) to allow the procedure to yield the desired precision.

THE PRIMARY PROCEDURE

With respect to generating the most information per animal, the LD50 is the most simple single summary statistic to measure on the dose-response curve. An up-and-down procedure is very efficient, in terms of the number of animals used, for obtaining this estimate. The up-and-down procedure specified in OECD TG 425 has been demonstrated to efficiently estimate the LD50, except when the step size is based on a "slope" estimate that is very far from reality or when the initial dose is distant from the LD50. A number of reasonable suggestions are made to mitigate these problems.

1. Stopping rule #3 involves those special cases when the procedure has not stopped at or before the nominal sample size is achieved. In this case, the recommendation is to stop if the likelihood ratio statistics for testing whether the true LD50 is 2.5 times greater than the estimate or 1/2.5 less than the estimate are both greater than 2.5. Simulations show this modification yields a great improvement in the estimates, particularly, when the slope is low or the initial treatment is far from the LD50. These ideas are strongly endorsed.

2. One modification to stopping rule #3 that warrants consideration is to calculate the likelihood using MIE of the dose-response function. MIEs have the advantage of (1) being very easy to calculate (a laboratory technician can compute MIEs without need of a computer; see Addendum IV of this report) and (2) not requiring an estimate of sigma when using the null hypothesis. An estimate of the slope is required for calculating the likelihood under the alternative hypotheses used in stopping rule #3.

3. Assuming a probit response function, a crude estimate of sigma can be obtained from the MIE of the dose-response function (rather than using a default estimate). Sigma can be estimated, for example, by noting that LD50-_ sigma is the 31st percentile of the normal probability density and LD50+_ sigma is the 69th percentile. Reading off the 31st and 69th percentiles (LD31 and LD69) of the interpolated isotonic estimate of the dose-response function, an estimate of sigma is (LD69-LD31). In addition, 2*(LD50-LD31) and 2*(LD69-LD50) provide two estimates of sigma. If they are very close to each other, the estimate (LD68-LD32)/2 should be

reasonable. A large difference might reflect the small sample size or it might indicate that the dose-response function is not symmetric, as is assumed by the probit model. Because of the relatively large interval between doses in the revised UDP Primary Test, it might be reasonable for the purpose of stopping to estimate sigma using estimates of LDp values that are more distant from the LD50 than are the LD31 and LD69 (e.g., LD16 and LD84). Because the data are clustered around the LD50, any estimate of sigma will not be very accurate, but it is worth evaluating whether this approach is better than assuming the default when the default is not true.

4. Future work, which should not interfere with the adoption of the current proposal, includes obtaining the exact distribution of the likelihood ratio statistics. This task will permit the critical value of 2.5 to be adjusted to satisfy the accuracy required for a particular application and should not be too difficult to accomplish assuming a (probit) model.

5. It needs to be emphasized that a variable stopping rule is essential in dose-response studies, because the investigator does not know how distant the initial dose level is from the LD50 (see Flournoy, 1993, for example). The development team for the revised UDP Primary Test recognized this need in developing the revised test.

6. Another recommendation is to increase the default step size. The recommendation is to adopt this proposal at this time. However, the issue of maintaining a constant step size throughout the experiment deserves additional investigation. For example, in the psychometrics literature (*cf.* Levitt, 1970), recommendations include doubling the step size after a string of like responses and halving the step size after a string of consecutive reversals. A procedure such as this could reduce the number of animals needed to get into the region of the LD50 (due to starting far away) and decrease the width of a confidence interval around the LD50 (when a steep dose-response curve causes many consecutive reversals).

Producing a reasonable algorithm for changing the step size is a considerable effort, in and of itself, and becomes even greater when the varied purposes for which this UDP is proposed are considered. Consequently, it is not recommended that this subject be investigated for the current proposal to OECD, but be included in future revisions.

MISCELLANEOUS DETAILS

The term “LD50” should not be used for both the parameter and the estimate. This wording is confusing in the BRD.

Also, there is an objection to a dose-escalation procedure being referred to as an up-and-down design. The up-and-down design with a nominal sample size of two is a simple dose-escalation procedure, as there is no decrease in exposure levels. It will have none of the nice features of the biased coin up-and-down design, such as clustering treatments around a target LDp. To refer to dose escalation as an up-and-down procedure is equivalent to treating all the animals at the same dose level, but stating that they were treated according to the normal probability density with variance equal to zero.

Addendum IV: Modified Isotonic Estimates of the Dose-Response Function

Reviews of isotonic estimation can be found in Barlow et al. (1972) and Robertson et al. (1988), among others. Modified isotonic estimates (MIE) of the dose-response curve were proposed by Stylianou (2000) and are reported in Stylianou and Flournoy (2000). A brief description is given below.

At each dose, the proportion of deaths observed is calculated. These proportions are reconsidered beginning at the lowest dose level. The proportion of animals that died at the lowest dose is the isotonic estimate of the probability of death at this dose. If the proportion of deaths at the next higher dose level is larger than the first proportion, it is the isotonic estimate of the probability of death at the second dose level. At successively higher doses, the proportion of animals that died is considered to be the isotonic estimate of the death rate, until a proportion is observed that is lower than the previous proportion. The dose-response function should increase with dose. When the data are inconsistent with this assumption, a weighted average of the two proportions is calculated, with weights equal to the sample sizes at the two dose levels. The weighted average replaces the observed proportions of animals that died as the isotonic estimators. The investigator continues to compare each observed proportion of animals dying at a particular dose level with the proportion at the preceding dose level and combining estimates when they fail to increase with increasing dose level. When the highest dose level has been considered, all of the isotonic estimates have been calculated.

The isotonic estimators are calculated only at the dose levels used in the experiment. An estimate of the death rate at any dose level is obtained by plotting the isotonic estimates and drawing lines between the points by hand or by computer. The curve that results from this linear interpolation is called the MIE and can be used with any acute toxicity procedure to estimate any LDp.

Up-and-down procedures cluster dose levels around target dose levels (see Addendum I of this report). If the up-and-down procedure in the revised UDP Primary Test is used, estimates of mortality at dose levels distant from the LD50 will not be very accurate; whereas, if a biased coin up-and-down procedure is used, the estimates will not be very accurate at dose levels distant from the targeted LDp. As a consequence, estimates of mortality for a specified dose level need to be generated using a procedure that is appropriate for a particular goal.

Up-and-Down Procedure (UDP)

Peer Panel Report

August 21, 2001 Meeting

1.0 INTRODUCTION

This report provides the conclusions and recommendations of an independent scientific peer review panel (Panel) evaluation of a revised version of the Up-and-Down Procedure (UDP) (July 2001). The Panel convened in a public teleconference meeting on August 21, 2001, at the National Institute of Environmental Health Science (NIEHS), Research Triangle Park, North Carolina, U.S. The Panel reviewed the following:

- The revised draft UDP, modified in response to recommendations from the July 2000 Panel meeting;
- A proposed procedure for calculating the confidence interval (CI) for the estimated LD50; and
- A software program to aid in establishing test doses, determining when to stop the test, estimating the LD50, and providing a CI for the LD50. (see **Appendix C**).

The meeting was organized by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). *Federal Register* notices relevant to the meeting include a Notice of Availability and Request for Comments (NIEHS, 2001a) and Notice and Agenda of Public Teleconference (NIEHS, 2001b).

The UDP was proposed by the U.S. Environmental Protection Agency (U.S. EPA) to ICCVAM in April 2000 as an alternate for the existing conventional LD50 test (EPA 870.1100, 1998; OECD TG 401, 1987) used to evaluate the acute oral toxicity of chemicals. A previous version of the draft UDP test guideline was reviewed by the UDP Peer Review Panel at a meeting on July 25, 2000 organized by the NICEATM and ICCVAM (Final Report Section I). The revised draft UDP reviewed on August 21, 2001 incorporated modifications made in response to the conclusions and recommendations of the Panel at the July 2000 meeting.

1.1 Objectives of the Peer Panel Evaluation

The Panel was charged with evaluating the following:

- the extent to which the revised draft UDP test guideline (July 12, 2001) addressed the Panel's recommendations at the July 25, 2000 Peer Review Panel meeting
- the appropriateness and adequacy of the proposed procedure for calculating a CI for the LD50; and
- the adequacy and consistency of the software program for use in the revised draft UDP test guideline.

1.2 Conduct of the Meeting and Reports

The UDP Peer Panel Review Meeting, which was open to the public, was conducted via teleconference on August 21, 2001 (**Appendix E-2**). The meeting began with an introduction including an overview of the ICCVAM Test Method Review Process. The Panel convened and evaluated the appropriateness and suitability of the further revised draft UDP test guideline, the approach for obtaining the CI, and the suitability of the software program. Following an opportunity for public comment, the Panel provided conclusions and adjourned. A written report, summarizing the discussions, recommendations, and conclusions from the teleconference, was provided to ICCVAM/NICEATM and is included in this final report (Final Report Section II).

2.0 REVISED DRAFT UP-AND-DOWN PROCUDURE TEST GUIDELINE

Based on the conclusions and recommendations of the Panel from their meeting in July 2000, the UDP Technical Task Force revised the test method guideline for the proposed UDP Primary and Limit Tests, deleted the UDP Supplemental Test, and included a procedure for calculating the CI for the estimated LD50. This revised draft UDP test guideline (GUIDELINE FOR THE TESTING OF CHEMICALS: Acute Oral Toxicity: Revised Up-and-Down Procedure. Draft, July 12, 2001; **Appendix C-1**) was developed by UDP Technical Task Force and submitted to ICCVAM on July 12, 2001. (Note: The slope of the dose-response curve was not addressed by the revised draft UDP test guideline.)

2.1 Panel Agreement on Guideline Revisions

The Panel concluded many of the recommended and requested changes had been appropriately considered and all members concurred with the current modifications. However, several previous recommendations appeared to have not been adequately addressed in the revised draft UDP test guideline, including the following:

- To increase flexibility and adaptability in animal use, the use of either sex or the more sensitive sex (if information is available indicating that one sex is more sensitive) should be permitted. The Panel unanimously re-affirmed this previous recommendation.

- The body weight of an animal on day 1 of dosing should be within 20% of the mean body weight of all previously dosed animals. The Panel chose to withdraw this recommendation based on the revised language included in paragraph 14 of the revised draft UDP test guideline as follows, "At the commencement of its dosing, each animal should be between 8 and 12 weeks old and its weight should fall in an interval ±20 % of the mean initial weight of all previously dosed animals" (**Appendix C-1)**.

- Additional guidance for use of pre-start data (data available before the acute toxicity test is conducted) to aid in determining the starting dose level should be included. The revised draft UDP test guideline addresses this recommendation in paragraph 4 as follows: "All available information on the test substance should be considered by the testing laboratory prior to conducting the study. Such information will include the identity and chemical structure of the substance; its physical chemical properties; the results of any other *in vitro* or *in vivo* toxicity tests on the substance or mixtures; toxicological data on structurally related substances or similar mixtures; and the anticipated use(s) of the substance. This information is useful to determine the relevance of the test for the protection of human health and the environment, and will help in the selection of an appropriate starting dose" (**Appendix C-1)**.

- Several Panel members stated this type of information was more appropriate for inclusion in a training session or guidance document, rather than a test guideline. The rationale for this recommendation was to help provide a better idea of the types of information or data to consider when selecting a starting dose level and to provide an alternative for the default starting dose level. The Panel unanimously recommended the following modification to the revised draft UDP test guideline, paragraph 4: All available information on the test substance should be considered by the testing laboratory prior to conducting the study. Such information <u>may</u> include the identity and chemical structure of the substance; its physical chemical properties; the results of any other *in vitro* or *in vivo* toxicity tests on the substance or mixtures; toxicological data on structurally related substances or similar mixtures; and the anticipated use(s) of the substance. This information is useful to determine the relevance of the test for the protection of human health and the environment. <u>This information may be valuable in selecting a dose other than the default starting dose</u>.

- The Panel unanimously re-affirmed their previous recommendation for a practicability evaluation of the revised UDP test guideline.

- A separate section in the revised UDP test guideline describing how the revised UDP Primary Test addresses reduction, refinement, and replacement of animals compared to the previous tests should be provided. The UDP Technical Task force formed the following response to this recommendation: *The Guideline significantly reduces the number of animals used in comparison to OECD Test Guideline 401, which often required at least 20 animals in a test: 1) the stopping rule limits the number of animals in a test; 2) sequential dosing introduces further efficiencies in animal use; 3) initial dosing is now set to be below the LD50, increasing the percentage of animals in which dosing levels will be sub lethal and thereby providing some reduction in pain and distress; and 4) the use of a single sex reduces the number of animals needed and minimizes the variability in the test population. Theoretically using females only could lead to an oversupply of males. However, the use of male rats in animal research greatly exceeds that of females and, thus, the preference for females in acute toxicity testing may well result in a better overall balance of the use of both genders. Importantly, the guideline contains a requirement to follow the OECD Guidance Document on the Recognition, Assessment, and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation (2000) that should reduce the overall suffering of animals used in this type of toxicity test.*

- The removal of gender specific references or the addition of the acceptability to use either gender (as per the preceding recommendation) was suggested and unanimously agreed upon by the Panel (see the underlined sentences in the above paragraph). This information should be included in the revised UDP test guideline.

- In paragraph 17a of the revised draft UDP test guideline, constant concentration should be used unless there is scientific or regulatory need for using constant volume. If constant volume is used in the performance of the UDP, concentrations used should also be provided. The Panel unanimously recommended that this statement be added to the revised UDP test guideline.

2.2 Recommendations

- The use of either sex of animals or the more sensitive sex (if information is available indicating one sex is more sensitive) should be permitted.
- Additional guidance pertaining to the use of pre-start data (data available before the acute toxicity test is conducted), which may be helpful in determining the starting dose level, should be provided.
- A practicability evaluation of the revised UDP test guideline should be conducted.
- A separate section detailing how the revised UDP Primary Test addresses reduction, refinement, and replacement of animals compared to the previous tests should be included.
- The Panel continues to express concerns that sufficient explanation is not included in the revised draft UDP test guideline describing the need and use of slope and CI for risk assessment and extrapolation to low doses for any purpose.

In addition to the above recommendations, the Panel identified the following editorial recommendations for the revised draft UDP test guideline:

- Check the text for the use of both “half-log unit” and “dose progression factor of 3.2” in the same sentence.
- Check whether the sentence in paragraph 10 should read “A test dose of 2000” rather than “A test dose of up to 2000”.
- Check for inconsistency in the number of stopping criteria. Annex 3 indicates four stopping criteria, but only three are described in the text.

- Check page 12 for the requirement of supplying a slope.
- Check to ascertain whether differences truly exist in the manner in which the 2000 mg/kg limit test is conducted compared to the 5000 mg/kg limit test. One test indicates dosing one animal at a time and the other indicates dosing in pairs. If the guideline is correct as written, a sentence concerning the rationale for the difference should be included.
- Check paragraph 27 and Annex 2 for consistency. Paragraph 27 suggests increasing the progression factor if the slope is <2.5. No recommendations are made for circumstances in which the slope >2.5, although Annex 2 details such cases. If smaller dose progression factors are recommended for steep slopes, a statement of this information should be included; otherwise, Annex 2 should be amended to accommodate only shallow slopes.
- Check paragraph 36 for clarity. Parts of paragraph 36 are unclear and the reference to paragraph 39 is not helpful. Perhaps a better explanation would be "An estimate of the log of the true LD50 is given by the value of mu (μ) to maximize the likelihood L."
- Clarify statements which include "OECD" (paragraphs 8, 38 and 40 for example). There is confusion about what the documents are called and how many exist.
- Include optional clinical chemistry in paragraph 34.
- Include an explanation for the use of 5 animals in the limit test.
- Check page 16, Stopping Rule. Consider including reference to both paragraphs 5 and 28.

3.0 CONFIDENCE INTERVAL PROCEDURE

Calculation of confidence intervals (CI) provides a basis for evaluating how to incorporate test results into regulatory applications. Therefore, a CI calculation was included in previous versions of the UDP guideline (OECD 1998a and ASTM 1998). Following deletion of the proposed supplemental procedure from the previous draft Revised UDP as per recommendation by the July 2000 Panel review, another method was needed to assist the investigator using the UDP to calculate a CI for the LD50. Based on this need, the U.S. EPA developed a proposed procedure for obtaining the CI; this procedure is a statistical calculation that does not require the use of test animals beyond what is needed to estimate the LD50 (**Appendix C-2)**. Further, the procedure helps to place the estimated LD50 in a statistical context for hazard and risk assessment purposes.

The UDP Panel charged Drs. Condon, Flournoy, and Stallard (the Panel's biostatisticians) with developing the Panel's position for this section by determining the appropriateness and adequacy of the procedure for calculating a CI for use with the revised draft UDP test guideline. It was recommended that language be added to the revised UDP test guideline to specifically indicate the shortcomings, uncertainties, and limitations of the CI procedure. Further, the procedure should be modified accordingly as more is learned about the use of these types of statistical methods.

3.1 Recommendations

1. Circumstances in which the proposed method does not perform well should be stated. The addition of non-statistical language and the outlining of specific situations in which the procedure does not perform well (e.g., shallow slopes) should be included in the revised UDP test guideline and the software program documentation. To aid in this task, appropriate references as suggested by the Panel included Jennison and Turnbull, 2000; Woodroofe, 1982; Liu, 1997; and Shiryaev and Spokoiny, 2000.

2. A very strong cautionary statement concerning the use of results for extrapolation to responses at lower dose levels is needed.

3. The fact that infinite confidence bounds can be obtained by this method should be stated.

4. A stronger cautionary statement pertaining to the utilization of a starting dose at the LD50 should be provided. If the LD50 is used as the starting dose level, a much wider confidence interval is obtained than if a higher or lower starting dose were used.

5. The revised UDP test guideline should state that evaluation of this method and examination of alternative approaches, such as nonparametric methods, should be encouraged.

4.0 SOFTWARE PROGRAM

A software program was designed and made publicly available to aid in the UDP test guideline procedures, to facilitate performance of the UDP, and to mitigate its complexity for the user (**Appendix C-3**). The U.S. EPA developed the "Acute Oral Toxicity (Guideline 425) Statistical Program" (AOT425StatPgm) to perform the statistical calculations associated with the OECD GUIDELINE FOR THE TESTING OF CHEMICALS, Section 4: Health Effects Test No. 425, Acute Oral Toxicity: Up-and-Down Procedure (OECD TG 425). The program may also be used with the revised draft UDP test guideline. The AOT425StatPgm program performs the calculations required to complete the test procedure by calculating 1) the doses for the test animals, 2) when to stop dosing animals, and 3) the specified LD50 and a confidence interval for the LD50. Additionally, the U.S. EPA conducted quality assurance testing and simulation testing to assess the performance of the software program and to determine the statistical performance of the OECD TG 425 procedure under various conditions.

With the charge of determining the sufficiency of the software, the Panel unanimously agreed that the software program to accompany the UDP is adequate and consistent with the procedures in the revised draft UDP test guideline. In the future, the program may need minor revisions as related to the evaluation of this method and examination of alternative approaches, such as nonparametric methods, as recommended in Section 3.1.

5.0 References

American Society for Testing and Materials (ASTM). 1998. Standard Test Method for Estimating Acute Oral Toxicity in Rats. ASTM E1163-98. In: Annual Book of ASTM Standards, Philadelphia.

Jennison, C. and B.W. Turnbull. 2000. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC: Boca Raton, FL.

Liu, A. 1997. On the MLE for the drift of Brownian motion following a symmetric SPRT. Comm. Statist.-Theory Meth. 26: 977-989.

National Institute of Environmental Health Sciences (NIEHS). 2001a. National Toxicology Program: The Revised Draft Up-and-Down Procedure for Assessing Acute Oral Toxicity. Notice of Availability and Request for Public Comments. 66 FR 33550. June 22, 2001.

NIEHS. 2001b. National Toxicology Program: The Revised Draft Up-and-Down Procedure for Assessing Acute Oral Toxicity. Notice of Peer Review Meeting. 66 FR 36294. July 11, 2001.

Organisation for Economic Co-operation and Development (OECD). 2000. Guidance Document on the Recognition, Assessment, and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation, OECD Environmental Health and Safety Publications, Series on Testing and Assessment, No. 19. OECD, Paris. 44 pp. Available: http://www.oecd.org//ehs/test/monos.htm.

OECD. 1998a. OECD Guideline for Testing Chemicals 425: Acute Oral Toxicity – Up-and-Down Procedure. OECD, Paris. [updated in 2000]. Available: http://www.oecd.org/ehs/test/health.htm.

OECD. 1987. OECD Guideline for Testing Chemicals 401: Acute Oral Toxicity. OECD, Paris.

Shiryaev, A.N and V.G. Spokoiny. 2000. Statistical Experiments and Decisions. Statistical inference for autoregressive models of the first order asymptotic theory. vol 8, chapter 5. World Scientific Publ., London, Singapore.

U.S. EPA. Health Effects Test Guidelines, OPPTS 870.1100, Acute Oral Toxicity. Washington, DC: U.S. Environmental Protection Agency, 1998. Available: http://www.epa.gov/docs/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/ [cited September 12, 2001].

Woodroofe, M. 1982. Nonlinear Renewal Theory in Sequential Analysis. BMS-NSF Regional Conference Series in Applied Mathematics 39, Section 1.3. SIAM.

III. REFERENCES

This reference list includes all cited references in the Peer Review Panel reports and key references provided to or recommended for the UDP Peer Review Panel.

16 CFR 1500. 2000. Title 16: Commercial Practices. Chapter II. Consumer Product Safety Commission. Part 1500: Hazardous Substances and Articles; Administration and Enforcement Regulations. Government Printing Office, Washington, DC.

29 CFR 1910.1200. 1998. Title 29: Department of Labor. Chapter XVII. Part 1910: Occupational Safety and Health Administration. Subpart Z: Toxic and Hazardous Substances. Section 1200: Hazard Communication. Government Printing Office, Washington, DC.

40 CFR 152. 2000. Title 40: Protection of Environment Agency. Code of Federal Regulations. Part 152: Pesticide Registration and Classification Procedures. Government Printing Office, Washington, DC.

40 CFR 156. 2000. Title 40: Protection of Environment Agency. Code of Federal Regulations. Part 156: Labeling Requirements for Pesticides and Devices. Government Printing Office, Washington, DC.

40 CFR 158. 2000. Title 40: Protection of Environment Agency. Code of Federal Regulations. Part 158: Data Requirements for Registration. Government Printing Office, Washington, DC.

40 CFR 721. 2000. Title 40: Protection of Environment Agency. Code of Federal Regulations. Part 721: Significant New Uses of Chemical Substances. Government Printing Office, Washington, DC.

49 CFR 173. 1999. Title 49: Department of Transportation. Code of Federal Regulations. Part 173: Shippers--General Requirements for Shipments and Packagings. Government Printing Office, Washington, DC.

American Society for Testing and Materials (ASTM). 1998. Standard Test Method for Estimating Acute Oral Toxicity in Rats. ASTM E1163-98. In: Annual Book of ASTM Standards, Philadelphia.

ASTM. 1987. Standard Test Method for Estimating Acute Oral Toxicity in Rats. ASTM E1163-87. In: Annual Book of ASTM Standards, Philadelphia.

BMDP Statistics Software, Inc. 1990. BMDP Statistical Software Manual. W.J. Dixon, Chief Ed. 1990 rev. or later. University of California Press, Berkeley, CA, USA.

Barlow, R.E., D.J. Bartholomew, J.M. Brenner, and H.D. Brunk. 1972. Statistical Inference Under Order Restrictions: The theory and application of isotonic regression. John Wiley & Sons, New York. 388 pp.

Bonnyns, E., M.P. Delcour, and A. Vral. 1988. Up-and-Down Method as an Alternative to the EC-Method for Acute Toxicity Testing. IHE Project No. 2153/88/11. Institute of Hygiene and Epidemiology, Ministry of Public Health and the Environment, Brussels. 33 pp.

Bruce, R.D. 1987. A Confirmatory Study for the Up-and-Down Method for Acute Toxicity Testing. Fundam. Appl. Toxicol. 8: 97-100.

Bruce, R.D. 1985. An Up-and-Down Procedure for Acute Toxicity Testing. Fundam. Appl. Toxicol. 5: 151-157.

Dixon, W.J. 1991. Design And Analysis of Quantal Dose-Response Experiments (with emphasis on staircase designs). Dixon Statistical Associates, Los Angeles CA, USA.

Dixon, W.J. 1991. Staircase Bioassay: The up-and-down method. Neurosci. Biobehav. Rev. 15:47-50.

Dixon, W.J. 1965. The Up-and-Down Method for Small Samples. J. Am. Stat. Assoc. 60:967-978.

Dixon, W.J. and A.M. Mood. 1948. A Method for Obtaining and Analyzing Sensitivity Data. J. Am. Stat. Assoc. 48:109-126.

Durham, S.D. and N. Flournoy. 1995. Up-and-Down Designs I: Stationary treatment distributions. In: Adaptive Designs; Flournoy, N. and W.F. Rosenberger (Eds.). Hayward, California: Institute of Mathematical Sciences. pp. 139-157.

Durham, S.D. and N. Flournoy. 1994. Random Walks for Quantile Estimation. In: Statistical Decision Theory and Related Topics V; Gupta, S.S. and J.O. Berger (Eds.). New York: Springer-Verlag. pp. 467-476.

Durham, S.D., N. Flournoy, and A.A. Montazer-Haghighi. 1995. Up-and-Down Designs II: Exact treatment moments. In: Adaptive Designs; Flournoy, N. and W.F. Rosenberger (Eds.). Hayward, California: Institute of Mathematical Sciences. pp. 158-178.

Durham, S.D., N. Flournoy, and W.F. Rosenberger. 1997. A Random Walk Rule for Phase I Clinical Trials. Biometrics 53:745-760.

Ekwall, B. 1999. Overview of the Final MEIC Results: II. The *in vitro*/*in vivo* evaluation, including the selection of a practical battery of cell tests for prediction of acute lethal blood concentrations in humans. Toxicol. *In Vitro* 13(4-5):665-673.

Finney, D.J. 1978. Statistical Methods in Biological Assay. Griffin, London.

Finney, D.J. 1971. Probit Analysis. 3rd ed. Cambridge Univ. Press, Cambridge, U.K.

Flournoy, N. 1993. A Clinical Experiment In Bone Marrow Transplantation: Estimating a percentage point of a quantal response curve. In: Case Studies in Bayesian Statistics; Gatsonis, C, J.S. Hodges, R.E. Kass, and N.D. Singpurwala (Eds.). New York: Springer-Verlag. pp.324-336.

Galson, S. 2000. Historical and Current Regulatory Perspectives. Opening Plenary Session, ICCVAM International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity, October 17-20, 2000.

Griffith, J.F. 1964. Interlaboratory Variations in the Determination of Acute Oral LD50. Toxicol. Appl. Pharmacol. 6: 726-730.

Gross, S.B. April 18, 1989. Memorandum Subject: Comments on standard evaluation procedure. Inhalation Toxicology Testing (SEP/Inhalation).

Gross, S.B. and F.J. Vocci. August 1988. Hazard Evaluation Division Standard Evaluation Procedure Inhalation Toxicity Testing. U.S. EPA-540/09-88-101.

Halle, W. 1998. Toxizitätsprüfungen In Zellkulturen Für Eine Vorhersage Der Akuten Toxizität (LD50) Zur Einsparung Von Tierversuchen. Life Sciences/ Lebens-wissenschaften, Volume 1, 94 pp., Jülich: Forschungszentrum Jülich.

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). 2001a. Guidance Document on Using *In Vitro* Data to Estimate *In Vivo* Starting Doses for Acute Toxicity. NIH Publication No. 01-4500. National Institute of Environmental Health Sciences. Research Triangle Park, North Carolina. Available: http://iccvam.niehs.nih.gov/methods/invidocs/guidance/iv_guide.pdf [cited October 18, 2001].

ICCVAM. 2001b. Report of the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity. NIH Publication No. 01-4499. National Institute of Environmental Health Sciences. Research Triangle Park, North Carolina. Available: http://iccvam.niehs.nih.gov/methods/invidocs/finalall.pdf [cited October 18, 2001].

ICCVAM. 2001c. The Revised Up-and-Down Procedure: A test method for determining the acute oral toxicity of chemicals. NIH Publication 02-4501. National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina.

ICCVAM. 2000. The Revised Up-and-Down Procedure: A test for determining the acute oral toxicity of chemicals and products. Proposed test method and background review document, April 14, 2000. National Institute of Environmental Health Sciences. Research Triangle Park, North Carolina. Available: http://iccvam.niehs.nih.gov/methods/udpdocs/AllBRDlk.pdf [cited October 18, 2001].

ICCVAM. 1999. Evaluation of the Validation Status of Toxicological Methods: General guidelines for submissions to ICCVAM validation and regulatory acceptance of toxicological test methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication 99-4496. National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina. Available: http://iccvam.niehs.nih.gov/docs/guidelines/subguide.htm [cited October 18, 2001].

ICCVAM. 1997. Validation and Regulatory Acceptance of Toxicological Test Methods: A report of the *ad hoc* Interagency Coordinating Committee on the Validation of Alternative Methods. NIH Publication 97-3981. National Institute of Environmental Health Sciences, Research Triangle Park, NC. Available: http://iccvam.niehs.nih.gov/docs/guidelines/validate.pdf [cited October 18, 2001].

Jaeger, B. November 1984. Pesticide Assessment Guidelines Subdivision F Hazard Evaluation: Human and Domestic Animals. Office of Pesticide Programs. U.S. EPA.

Jennison, C. and B.W. Turnbull. 2000. Group Sequential Methods with Applications to Clinical Trials. Chapman & Hall/CRC: Boca Raton, FL.

Levitt, H. 1971. Transformed Up-Down Methods in Psychoacoustics. J. Acoustical Soc. America 49:467-447.

Lipnick, R.L., J.A. Cotruvo, R.N. Hill, R.D. Bruce, K.A. Stitzel, A.P. Walker, I. Chu, M. Goddard, L. Segal, J.A. Springer, and R.C. Myers. 1995. Comparison of the Up-And-Down, Conventional LD50, and Fixed-Dose Acute Toxicity Procedures. Food Chem. Toxicol. 33: 223-231.

Litchfield, J.T., Jr. and F. Wilcoxon. 1949. A Simplified Method of Evaluating Dose-Effect Experiments. J. Pharmacol. Exp. Ther. 96:99-113.

Liu, A. 1997. On the MLE for the Drift of Brownian Motion Following a Symmetric SPRT. Comm. Statist.-Theory Meth. 26: 977-989.

Luttik, R. and T. Aldenberg. Extrapolation Factors for Small Samples of Pesticide Toxicity Data: Special focus on LD50 values for birds and mammals. Environ. Toxicol. Chem. 16:1785-1788.

Mats, V.A., W.F. Rosenberger, and N. Flournoy. 1998. Restricted Optimality for Phase I Clinical Trials. In: New Developments and Applications in Experimental Designs; Flournoy, N., W.F. Rosenberger, and W.K. Wong (Eds.). IMS Monograph Series 34:50-61.

Miller, F.J. (Ed.). 1994. Nasal Toxicity and Dosimetry of Inhaled Xenobiotics, Implications for Human Health. Chemical Industry Institute of Toxicology. Taylor & Francis, Washington, D.C. pp. 452-455.

Mulder, G.J. 1986. Sex Differences in Drug Conjugation and Their Consequences for Drug Toxicity. Sulfation, glucuronidation and glutathione conjugation. Chem. Biol. Interactions 57:1-15.

National Institute of Environmental Health Sciences (NIEHS). 2000a. National Toxicology Program: Request for Data and Nomination of Expert Scientists to Participate in the Independent Peer Review Evaluation of the Revised Up-and-Down Procedure for Assessing Acute Oral Toxicity. Evaluation of the Up-and-Down Procedure. 65 FR 8385. February 18, 2000.

NIEHS. 2000b. National Toxicology Program: Notice of Peer Review Meeting on the Revised Up-and-Down Procedure (UDP) as an Alternative Test Method for Assessing Acute Oral Toxicity. Request for Comments. 65 FR 35109. June 1, 2000.

NIEHS. 2001a. National Toxicology Program: The Revised Draft Up-and-Down Procedure for Assessing Acute Oral Toxicity. Notice of Availability and Request for Public Comments. 66 FR 33550. June 22, 2001.

NIEHS. 2001b. National Toxicology Program: The Revised Draft Up-and-Down Procedure for Assessing Acute Oral Toxicity. Notice of Peer Review Meeting. 66 FR 36294. July 11, 2001.

Nelson, D.R., L. Koymans, T. Kamatski, J.J. Stegeman, R. Feyereisen, D.J. Waxman, M.R. Waterman, O. Gotoh, M.J. Coon, R.W. Estrabrook, I.C. Gunsalus, and D.W. Nebert. 1996. P450 Super Family: Update on new sequences, gene mapping accession numbers and nomenclature. Pharmacogenetics 6:1-42.

Organisation for Economic Co-operation and Development (OECD). 2001. Harmonised Integrated Classification System for Human Health and Environmental Hazards of Chemical Substances and Mixture. OECD Chemicals Committee and the Working Party on Chemicals, Pesticides, and Biotechnology, Series on Testing and Assessment, No. 33. OECD, Paris. 247 pp. Available: http://www.oecd.org/ehs/class/HCL6.htm. [cited October 18, 2001].

OECD. 2000a. Guidance Document on the Recognition, Assessment, and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation, OECD Environmental Health and Safety Publications, Series on Testing and Assessment, No. 19. OECD, Paris. 44 pp. Available: http://www.oecd.org//ehs/test/monos.htm. [cited October 18, 2001].

OECD. 2000b. OECD Guideline for the Testing of Chemicals, Revised Draft Test Guideline 420, Acute Oral Toxicity - Fixed Dose Procedure. OECD, Paris. Available: http://www.oecd.org/ehs/test/health.htm [cited September 12, 2001].

OECD. 2000c. OECD Guideline for the Testing of Chemicals, Revised Draft Test Guideline 423, Acute Oral Toxicity - Acute Toxic Class Method. OECD, Paris. Available: http://www.oecd.org/ehs/test/health.htm [cited September 12, 2001].

OECD. 2000d. OECD Guideline for the Testing of Chemicals, Revised Draft Test Guideline 425, Acute Oral Toxicity - Up-and-Down Procedure. OECD, Paris. Available: http://www.oecd.org/ehs/test/health.htm [cited September 12, 2001].

OECD. 1998a. OECD Guideline for the Testing of Chemicals, Test Guideline 425, Acute Oral Toxicity - Up-and-Down Procedure. OECD, Paris. Available: http://www.oecd.org/ehs/test/health.htm [cited September 12, 2001].

OECD. 1998b. Harmonized Integrated Hazard Classification System for Human Health and Environmental Effects of Chemical Substances as Endorsed by the 28th Joint Meeting of the Chemicals Committee and the Working Party on Chemicals in November 1998, Part 2, p. 11. Available: http://www.oecd.org/ehs/class/HCL6.htm. [updated August 14, 2001 -- see also OECD, 2001].

OECD. 1996. OECD Guideline for the Testing of Chemicals, Test Guideline 423, Acute Oral Toxicity - Acute Toxic Class Method. OECD, Paris. Available: http://www.oecd.org/ehs/test/health.htm [cited September 12, 2001].

OECD. 1992. OECD Guideline for the Testing of Chemicals, Test Guideline 420, Acute Oral Toxicity - Fixed Dose Procedure. OECD, Paris. Available: http://www.oecd.org/ehs/test/health.htm [cited September 12, 2001].

OECD. 1987. OECD Updated Guideline for the Testing of Chemicals Test Guideline 401, Acute Oral Toxicity. OECD, Paris. Available: http://www.oecd.org/ehs/test/health.htm [cited September 12, 2001].

OECD. 1981. OECD Guideline for the Testing of Chemicals 401: Acute Oral Toxicity. OECD, Paris.

Robertson, T., F.T. Wright, and R.L. Dykstra. 1988. Order Restricted Statistical Inference, John Wiley & Sons, New York.

SAS Institute Inc. 1990. SAS/STAT® User's Guide. Version 6, Fourth Ed. or later. Cary, NC, USA.

Salem, H. 1987. Inhalation Toxicology. Research Methods, Applications, and Evaluation. Marcel Dekker, Inc., New York.

Schlede, E., W. Diener, U. Mischke, and D. Kayser. 1994. OECD Expert Meeting: Acute toxic class method. January 26-28, 1994, Berlin, Germany.

Schlede, E., U. Mischke, W. Diener, and D. Kayser. 1995. The International Validation Study of the Acute Toxic Class Method (oral). Arch. Toxicol. 69: 659-670.

Schlede, E., U. Mischke, R. Roll, and D. Kayser. 1992. A National Validation Study of the Acute Toxic Class Method - An alternative to the LD50 test. Arch. Toxicol. 66: 455-470.

Sheehan P.J., A. Baril, P. Mineau and D. Paustenbach. 1995. Predicting the Effects of Pesticides on Aquatic Systems and the Waterfowl that Use Them. Chapter 30 in Fundamentals of Aquatic Toxicology. Second edition, G.M. Rand (Ed.), Taylor and Francis, North Palm Beach, Florida. pp. 827-857.

Shiryaev, A.N and V.G. Spokoiny. 2000. Statistical Experiments and Decisions. Statistical inference for autoregressive models of the first order asymptotic theory. Vol 8, chapter 5. World Scientific Publ., London, Singapore.

Sitter, R.R. and C.F.J. Wu. 1993. Optimal Designs for Binary Response Experiments - Fieller, D, and A criteria. Scandinavian J. Statistics 20:329-341.

Spielmann, H., E. Genschow, M. Liebsch, and W. Halle. 1999. Determination of the Starting Dose for Acute Oral Toxicity (LD50) Testing in the Up-and-Down Procedure (UDP) from Cytotoxicity Data. ATLA 27:957-966.

Stylianou, M. 2000. A New Approach to Dose Finding for Phase I Clinical Trials. Dissertation. American University.

Stylianou, M. and N. Flournoy. 2000. A New Approach to Dose Finding for Phase I Clinical Trials. Technical Report Number 2000-2. Department of Mathematics and Statistics. American University.

Technical Committee of the Inhalation Specialty Section. 1992. Society of Toxicology Recommendations for the Conduct of Acute Inhalation Limit Tests. Fundam. Appl. Toxicol. 18: 321-327.

Trevan, J.W. 1927. The Error of Determination of Toxicity. Proc. Royal Soc. 101B: 483-514.

U.S. EPA. Health Effects Test Guidelines, OPPTS 870.1100, Acute Oral Toxicity. Washington, DC: U.S. Environmental Protection Agency, 1998. Available: http://www.epa.gov/docs/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/ [cited October 18, 2001].

U.S. EPA. Health Effects Test Guidelines, OPPTS 870.1300, Acute Inhalation Toxicity. Washington, DC: U.S. Environmental Protection Agency, 1998. Available: http://www.epa.gov/docs/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/ [cited October 18, 2001].

van den Heuvel, M.J., D.G. Clark, R.J. Fielder, P.P. Koundakjian, G.J.A. Oliver, D. Pelling, N.J. Tomlinson, and A.P. Walker. 1990. The International Validation of a Fixed-Dose Procedure as an Alternative to the Classical LD50 Test. Food Chem. Toxicol. 28: 469-482.

van den Heuvel, M.J., A.D. Dayan, and R.O. Shillaker. 1987. Evaluation of the BTS Approach to the Testing of Substances and Preparations for Their Acute Toxicity. Human Toxicol. 6: 279- 291.

Weil, C.S. 1983. Economical LD50 and Slope Determinations. Drug Chem. Toxicol. 6:595-603.

Weil, C.S. 1975. Toxicology Experimental Design and Conduct as Measured by Interlaboratory Collaborative Studies. J. Off. Anal. Chem. 58: 683-688.

Weil, C.S. 1952. Tables for Convenient Calculation of Median-Effective Dose [LD50 or ED50] and Instructions in Their Use. Biometrics 8:249-263.

Weil, C.S., C.P. Carpenter, and H.F. Smyth. 1953. The Median Effective Dose. Ind. Hyg. Q. 14: 200-206.

Weil, C.S., C.P. Carpenter, J.S. West, and H.F. Smyth. 1966. Reproducibility of Single Oral Dose Toxicity Testing. Am. Ind. Hyg. Assoc. J. 27: 483-487.

Weil, C.S. and G.J. Wright. 1967. Intra- and Interlaboratory Comparative Evaluation of a Single Oral Test. Toxicol. Appl. Pharm. 11: 378-388.

Whalan, J.E. and J.C. Redden. Interim U.S. EPA Policy for Particle Size and Limit Concentration Issues in Inhalation Toxicity Studies.

Woodroofe, M. 1982. Nonlinear Renewal Theory in Sequential Analysis. BMS-NSF Regional Conference Series in Applied Mathematics 39, Section 1.3. SIAM.

Yam, J., P.J. Reer, and R.D. Bruce. 1991. Comparison of the Up-and-Down Method and the Fixed-Dose Procedure for Acute Oral Toxicity Testing. Food Chem. Toxicol. 29:259-263.

Zbinden, G. and M. Flury-Roversi. 1981. Significance of the LD50 Test for the Toxicological Evaluation of Chemical Substances. Arch Toxicol. 47: 77-99.

ICCVAM Test Recommendation for the Up-and-Down Procedure: A Test Method for Determining the Acute Oral Toxicity of Chemicals

I. Introduction

In August 1999, the U.S. Environmental Protection Agency (EPA) requested the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) to conduct an independent scientific peer review evaluation of the validation status of a revised Up-and-Down Procedure (UDP) for determining the acute oral toxicity of chemicals. The revised UDP was proposed as an alternative to the existing conventional LD50 test [OECD Test Guideline (TG) 401, 1987; U.S. EPA 870.1100, 1998]. An earlier version of the UDP test method had been adopted by the Organisation for Economic Co-operation and Development (OECD) Test Guidelines Program in 1998 (TG 425; OECD 1998a). The U.S. EPA subsequently determined it was necessary to revise the UDP. The revisions were needed to 1) conform to a newly harmonised global hazard classification scheme for acute toxicity (OECD, 1998b); and 2) to incorporate changes to ensure the regulatory and testing needs would be met using the revised UDP prior to the OECD's proposed deletion of the TG 401 (OECD, 1987).

The revised UDP test method submitted to ICCVAM included three components:

- A Primary Test, which provided an improved estimate of acute oral toxicity with an accompanying reduction in the number of animals used, when compared to TG 401 and the existing TG 425;
- A Limit Test for substances anticipated to have minimal toxicity; and
- A Supplemental Test to determine the slope and confidence interval (CI) for the dose-response curve.

II. ICCVAM Independent Scientific Peer Review

July 25, 2000 Peer Review Meeting

In a public session on July 25, 2000, ICCVAM convened an international independent scientific peer review panel (Panel) to evaluate the validation status of the revised UDP (Federal Register, NIEHS, 2000a, 2000b). The Panel evaluated the extent to which established validation and acceptance criteria (ICCVAM, 1997) had been addressed, and developed conclusions regarding the usefulness and limitations of the revised UDP. The Panel also responded to the following questions:

- Has the revised UDP been evaluated sufficiently, and is its performance satisfactory to support its adoption as a substitute for the currently accepted UDP (OECD, 1998a), and as a substitute for the conventional LD50 test for acute oral toxicity (U.S. EPA OPPTS 870.1100, 1998; OECD, 1987)?

- With respect to animal welfare, does the revised UDP adequately consider and incorporate where scientifically feasible, procedures to refine, reduce, and/or replace animal use?

The Panel's report is included in the publication: "The Revised Up-and-Down Procedure: A Test Method for Determining the Acute Oral Toxicity of Chemicals," NIH Publication 02-4501 (ICCVAM, 2001b). The Panel's conclusions were as follows:

- **UDP Primary Test**
  "The performance of the revised UDP Primary Test is satisfactory and exceeds the performance of OECD TG 401 in providing, with fewer animals, both an improved estimate of the LD50 for the purpose of hazard classification and more accurate information on acute toxicity. In particular, the use of 0.5 log units for dose spacing is reasonable and appropriate based on experience and the results of computer simulations. Three disadvantages of the revised UDP Primary Test recognized by the Panel are: a) the increased length of time needed to conduct a study; b) the increased costs per test material evaluated; and c) the increased complexity of the protocol."

- **UDP Limit Test**
  "The revised UDP Limit Test at 2000 or 5000 mg/kg is expected to perform as well as or better than the Limit Test in OECD TG 401, with a reduction in the number of animals needed to conduct a test."

- **UDP Supplemental Test**
  "The UDP Supplemental Test for slope and CI is not recommended for adoption. The Panel was unable to evaluate the utility of the test because sufficient information regarding the use of the resulting data was not provided. As a consequence, any impact on animal use was not assessed."

- **Animal Welfare Considerations**
  "The revised UDP Primary Test and the revised UDP Limit Test will reduce the number of animals used, but will not replace the use of animals. The Panel could not reach a consensus on the overall issue of refinement. However, the OECD Guidance Document on the Recognition, Assessment, and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluations (OECD, 2000), referenced in the revised UDP Guideline, provides an element of refinement."

<u>Revisions to the UDP in response to the July 25, 2000 Panel Report</u>

Based on the Panel's conclusions and recommendations from July 25, 2000, the UDP Technical Task Force further revised the UDP test method guideline as follows:

- Revisions recommended by the Panel were incorporated into the proposed UDP Primary and Limit Tests;
- The UDP Supplemental Test to determine the slope of the dose-response curve was deleted;
- A procedure was added (for use with the Primary Test) to calculate the CI for the estimated LD50. This procedure is a statistical method that does not require the use of additional

animals. The CI helps to place the estimated LD50 in a statistical context for hazard and risk assessment purposes.

- The U.S. EPA developed a publicly available software program for use in establishing test doses, determining when to stop the test, estimating the LD50, and providing a CI for the LD50.

The Technical Task Force also responded with the following clarifications regarding animal welfare:

- The revised UDP significantly reduces the number of animals used in comparison to OECD TG 401 by the incorporation of the following: 1) a stopping rule which limits the maximum number of animals in a test; and 2) a sequential dosing method which introduces greater efficiencies in animal use.
- The revised UDP provision that the initial starting dose should be below the LD50 will result in fewer animals receiving a lethal dose, thereby potentially providing further reduction in pain and distress.
- Adherence to the OECD Guidance Document on Humane Endpoints (2000) should provide additional reduction or minimization of pain and distress in animals used in this procedure.

The revised version of the UDP and the UDP software program were then provided to the Panel and made available for public comment in July 2001.

August 21, 2001 Peer Review Panel Meeting

The UDP Panel met, via public teleconference, on August 21, 2001 (Federal Register, NIEHS, 2001) to evaluate the appropriateness and suitability of the further revised UDP, the approach for obtaining the CI, and the suitability of the software program. Their conclusions and recommendations were as follows:

- **Further Revisions to the Revised UDP**
  The Panel concluded the changes made in the revised UDP Test Guideline were acceptable, but also recommended further clarifications to the UDP as follows:

  - Either sex of animal can be used, or if information is available indicating that one sex is more sensitive, the more sensitive sex should be used.
  - A practicability evaluation of the usability of the *in vivo* test should be conducted to supplement the computational analyses.
  - A separate section on how the revised UDP Primary Test addresses reduction, refinement, and replacement of animals when compared to the previous tests should be included to the UDP guideline.
  - Constant concentration in dosing should be used unless there is a clear scientific or regulatory justification for using constant volume. In the event that constant volume is used, information on the actual concentrations utilized should be provided.

- **CI Procedure**
  The Panel endorsed the proposed procedure for calculating the CI for the estimated LD50. However, the Panel recommended the inclusion of language in the UDP guideline and

software to fully describe the limitations and uncertainties of the proposed method, and to provide appropriate cautions for interpretation of test results. The Panel noted that statistical techniques are evolving and recommended the future development of alternative approaches, such as nonparametric methods, be encouraged.

· **Software Program**
The Panel concluded the software program was appropriate and suitable for establishing test doses, determining when to stop the test, estimating the LD50, and providing a CI for the LD50.

Revisions to the UDP in response to the August 21, 2001 Panel Report
Following the August 21, 2001 Panel meeting, the UDP Technical Task Force revised the UDP Guideline in response to the Panel's recommendations. This revised version is included in the Final Peer Review Report (ICCVAM, 2001b) as Appendix B.

III. ICCVAM Test Method Recommendations

The ICCVAM agrees with the UDP Peer Review Panel that the revised UDP test guideline, with incorporation of the Panel's recommendations from the August 21st Panel Meeting, is acceptable as a substitute for the conventional LD50 test for acute oral toxicity (U.S. EPA OPPTS 870.1100, 1998; OECD, 1987) for the purpose of hazard classification and for obtaining certain information on acute toxicity. The ICCVAM also agrees with the Panel that the revised UDP Test Guideline will reduce and refine animal use. The ICCVAM concurs also with the other conclusions and recommendations of the Panel.

ICCVAM therefore recommends that the final revised UDP test guideline should: (1) replace the current OECD UDP test guideline (TG 425; OECD, 1998a) and (2) be used instead of the conventional LD50 test to determine the acute oral toxicity hazard of chemicals.

The ICCVAM also concludes:

1. The revised UDP performs appropriately and will result in a reduction in animal usage compared to the conventional LD50 test. The recommendation to use a starting dose level below the anticipated LD50 and to follow the OECD Guidance Document on Humane Endpoints (2000) will refine animal use by decreasing pain and distress.

2. The revised UDP is an appropriate method for generating a point estimate for the LD50 for use in hazard classification and in estimating a CI for the LD50 under specified circumstances. The revised UDP does not provide information about the slope of the dose-response curve for lethality. If other human health or ecological risk assessment information is desired, including hazard dose-response and slope information, a different test should be conducted.

3. Compared to the conventional LD50 procedure, the UDP will require additional time. However, it provides potential improvements in animal welfare and is the only alternative

to OECD TG 401 that will generate a point estimate for the LD50 and an accompanying CI.

4. Compared to the conventional LD50 procedure, the UDP is computationally more complex. However, the UDP does provide increased statistical power with the use of sequential dosing. The publicly available statistical software will greatly simplify and facilitate efficient conduct of the UDP. The software calculates subsequent test dose levels, determines when testing is complete, estimates the LD50, and provides an appropriate and useful CI for the LD50.

5. Due to the reduction in the number of animals required when compared to the conventional LD50 test, the amount of test material needed will also be decreased.

6. The UDP may not be appropriate for chemicals causing delayed deaths (especially after five days).

7. Limit dose testing may be conducted at 2000 or 5000 mg/kg, depending on regulatory program needs.

8. For scientific purposes, the testing of three to five animals in the Limit Test is adequate. However, it is recognized that OECD stipulates utilizing five animals at 2000 mg/kg for all alternative acute toxicity methods as a way of harmonizing procedures.

9. Either sex can be used for the UDP. However, in the absence of information indicating males may be more sensitive, it is recommended that females be used based on available data showing females to be generally more sensitive.

10. Statistical methods are evolving rapidly, thereby providing reason to consider revisiting the UDP test design in the future.

11. A practicability assessment of the revised UDP should be considered.

12. *In vitro* data may be helpful in estimating an appropriate starting dose level for UDP studies. This approach may further reduce the number of animals needed, especially if the results indicate a Limit Test may be appropriate. Further guidance can be found in the "Guidance Document on Using *In Vitro* Data To Estimate *In Vivo* Starting Doses for Acute Toxicity", NIH Publication 01-4500 (ICCVAM, 2001a).

Adopted by ICCVAM:
October 10, 2001

References

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM). 1997. Validation and regulatory acceptance of toxicological test methods: A report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods. NIH publication no: 97-3981. National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina. Available: http://iccvam.niehs.nih.gov/docs/guidelines/validate.pdf [cited October 1, 2001].

ICCVAM. 2001a. Guidance Document on Using *In Vitro* Data to Estimate *In Vivo* Starting Doses for Acute Toxicity. NIH Publication 01-4500. National Institute of Environmental Health Research Triangle Park, North Carolina. Available: http://iccvam.niehs.nih.gov/methods/invidocs/guidance/iv_guide.htm [cited October 1, 2001].

ICCVAM. 2001b. The Revised Up-and-Down Procedure: A test method for determining the acute oral toxicity of chemicals. NIH Publication 02-4501. National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina.

National Institute of Environmental Health Sciences (NIEHS). 2000a. National Toxicology Program: Request for Data and Nomination of Expert Scientists to Participate in the Independent Peer Review Evaluation of the Revised Up-and-Down Procedure for Assessing Acute Oral Toxicity. Evaluation of the Up-and-Down Procedure. 65 FR 8385. February 18, 2000.

NIEHS. 2000b. National Toxicology Program: Notice of Peer Review Meeting on the Revised Up-and-Down Procedure (UDP) as an Alternative Test Method for Assessing Acute Oral Toxicity. Request for Comments. 65 FR 35109. June 1, 2000.

NIEHS. 2001. National Toxicology Program: The Revised Draft Up-and-Down Procedure for Assessing Acute Oral Toxicity. Notice of Availability and Request for Public Comments. 66 FR 33550. June 22, 2001.

Organisation for Economic Co-operation and Development (OECD). 2000. Guidance Document on the Recognition, Assessment, and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation, OECD Environmental Health and Safety Publications, Series on Testing and Assessment, No. 19. 44 pp. OECD, Paris. Available: http://www.oecd.org//ehs/test/monos.htm [cited October 1, 2001].

OECD. 1998a. OECD Guideline for the Testing of Chemicals, Revised Test Guideline 425, Acute Oral Toxicity, Up-and-Down Procedure. OECD, Paris. Available: http://www.oecd.org//ehs/test/health.htm [cited October 1, 2001].

OECD. 1998b. Harmonized Integrated Hazard Classification System for Human Health and Environmental Effects of Chemical Substances as endorsed by the 28th Joint Meeting of the Chemicals Committee and Working Party on Chemicals in November 1998. OECD, Paris. Available: http://www.oecd.org/ehs/class/HCL6.htm [cited October 1, 2001].

OECD. 1987. OECD Guideline for the Testing of Chemicals, Test Guideline 401, Acute Oral Toxicity. OECD, Paris. Available: http://www.oecd.org//ehs/test/health.htm [cited October 1, 2001].

U.S. EPA. 1998. Health Effects Test Guidelines, OPPTS 870.1100, Acute Oral Toxicity. Washington, DC, U.S. Environmental Protection Agency. Available: http://www.epa.gov/docs/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/ [cited October 1, 2001].

udpfin7
10/31/01

Acute Oral Toxicity: Up-and-Down Procedure

INTRODUCTION

1. The concept of the up-and-down testing approach was first described by Dixon and Mood (1)(2)(3)(4). In 1985, Bruce proposed to use an Up-and-Down Procedure (UDP) for the determination of acute toxicity of chemicals (5). There exist several variations of the up-and-down experimental design for estimating an LD50. This guideline is based on the procedure of Bruce as adopted by the American Society for Testing and Materials (ASTM) in 1987 (6) and revised in 1990. A study comparing the results obtained with the UDP, the conventional LD50 test and the Fixed Dose Procedure (FDP, Guideline 420) was published in 1995 (7). Since the early papers of Dixon and Mood, papers have continued to appear in the biometrical and applied literature, examining the best conditions for use of the approach (8)(9)(10)(11).

2. The test procedure described in this guideline is of value in minimizing the number of animals required to estimate the acute oral toxicity of a chemical. In addition to the estimation of LD50 and confidence intervals (CI), the test allows the observation of signs of toxicity. This test does not provide information about the slope of the dose-response curve. Appendix A contains definitions of some terms used in this guideline. Revision of this test guideline was undertaken concurrently with two other alternatives to conventional acute oral toxicity testing. An international guidance document on acute toxicity gives more information (12).

3. The guideline significantly reduces the number of animals used in comparison to Guideline 401, which often required at least 20 animals in a test: 1) the stopping rule limits the number of animals in a test; 2) sequential dosing introduces further efficiencies in animal use; 3) initial dosing is now set to be below the LD50 increasing the percentage of animals in which dosing levels will be sub lethal and thereby providing some reduction in pain and distress; and 4) the use of a single sex reduces the number of animals needed and minimizes the variability in the test population. Importantly, the guideline contains a requirement to follow the Organization for Economic Cooperation and Development (OECD) Guidance Document on Humane Endpoints (13) that should reduce the overall suffering of animals used in this type of toxicity test.

INITIAL CONSIDERATIONS

4. All available information on the test substance should be considered by the testing laboratory prior to conducting the study. Such information may include the identity and chemical structure of the substance; its physical chemical properties; the results of any other

in vitro or *in vivo* toxicity tests on the substance or mixtures; toxicological data on structurally related substances or similar mixtures; and the anticipated use(s) of the substance. This information is useful to determine the relevance of the test for the protection of human health and the environment. This information may be valuable in selecting a dose other than the default starting dose. (See caveats in paragraph 5 about placement of starting dose.) For example, data from *in vitro* cytotoxicity assays can be useful as one of the tools in setting a starting dose for the *in vivo* assessment of acute oral toxicity (14, 15, 16). A Guidance Document on Using *In Vitro* Data to Estimate *In Vivo* Starting Doses for Acute Toxicity is available (15), and preliminary information suggests that the use of this approach may further reduce the number of animals used for *in vivo* testing (16).

5. If no information is available to make a preliminary estimate of the LD50 and the slope of the dose-response curve, results of computer simulations have suggested that starting near 175 mg/kg and using half-log units (corresponding to a dose progression of 3.2) between doses will produce the best results. This starting dose should be modified if the substance is likely to be highly toxic. The half-log spacing provides for a more efficient use of animals, and increases accuracy in the prediction of the LD50 value. Because the method has a bias toward the starting dose, it is essential that initial dosing occur below the estimated LD50. (See paragraph 30 and Appendix B for discussion of dose sequences and starting values.) However, for chemicals with large variability (i.e., shallow dose-response slopes), bias can still be introduced in the lethality estimates and the LD50 will have a large statistical error, similar to other acute toxicity methods. To correct for this, the main test includes a stopping rule keyed to properties of the estimate rather than a fixed number of test observations. (See paragraph 31.)

6. The method is easiest to apply to materials that produce death within one or two days. The method would not be practical to use when considerably delayed death (five days or more) can be expected.

7. Computers are used to facilitate animal-by-animal calculations that establish testing sequences and provide final estimates.

8. Moribund animals or animals obviously in pain or showing signs of severe and enduring distress shall be humanely killed, and are considered in the interpretation of the test results in the same way as animals that died on test. Criteria for making the decision to kill moribund or severely suffering animals, and guidance on the recognition of predictable or impending death are the subject of a separate OECD guidance document (13).

9. A limit test can be used efficiently to identify chemicals that are likely to have low toxicity.

PRINCIPLE OF THE LIMIT TEST

10. The Limit Test is a sequential test that uses a maximum of 5 animals. A test dose of

2000, or exceptionally 5000 mg/kg, may be used. The procedures for testing at 2000 and 5000 mg/kg are slightly different. (See paragraphs 22 - 24 for limit test at 2000 mg/kg and paragraphs 25 - 28 for limit test at 5000 mg/kg.) The selection of a sequential test plan increases the statistical power and also has been made to intentionally bias the procedure towards rejection of the limit test for compounds with LD50s near the limit dose; i.e., to err on the side of safety. As with any limit test protocol, the probability of correctly classifying a compound will decrease as the actual LD50 more nearly resembles the limit dose.

PRINCIPLE OF THE MAIN TEST

11. The main test consists of a single ordered dose progression in which animals are dosed, one at a time, at 48-hour intervals. The first animal receives a dose a step below the level of the best estimate of the LD50. If the animal survives, the dose for the next animal is increased to a factor of one half log times the original dose; if it dies, the dose for the next animal is decreased by a similar dose progression. (Note: 3.2 is the default factor corresponding to a dose progression of one half log unit. Paragraph 30 provides further guidance for choice of dose spacing factor.) Each animal should be observed carefully for up to 48 hours before making a decision on whether and how much to dose the next animal. That decision is based on the 48-hour survival pattern of all the animals up to that time. (See paragraphs 29 and 33 on choice of survival interval.) A combination of stopping criteria is used to keep the number of animals low while adjusting the dosing pattern to reduce the effect of a poor starting value or low slope (see paragraph 32). Dosing is stopped when one of these criteria is satisfied (see paragraphs 31 and 39), at which time an estimate of the LD50 and a CI are calculated for the test based on the status of all the animals at termination. For most applications, testing will be completed with only 4 animals after initial reversal in animal outcome. The LD50 is calculated using the method of maximum likelihood (17)(18). (See paragraphs 39 and 41.)

12. The results of the main test procedure serve as the starting point for a computational procedure to provide a CI estimate where feasible. A description of the basis for this CI is outlined in paragraph 43.

DESCRIPTION OF THE METHOD

Selection of animals species

13. The preferred rodent species is the rat although other rodent species may be used. The test is conducted using a single sex in order to reduce variability and as a means of minimizing the number of animals used. Either sex may be used, however, if there is information available indicating differences in sensitivity, the most sensitive sex should be tested (12). Literature surveys of conventional LD 50 tests show that usually there is little difference in sensitivity between the sexes but, in those cases where differences were observed, females were often slightly more sensitive (7). For chemicals that are direct acting in their toxic mechanism, female

rats may have a lower detoxification capacity than males, as measured by specific activity of phase I and II enzymes. However, all available information should be evaluated, for example on chemical analogues and the results of testing for other toxicological endpoints on the chemical itself, as this may indicate that males may be more sensitive than females. Knowledge that metabolic activation is required for a chemical's toxicity can also indicate that males may be the more sensitive sex.

Occasionally, the results of subsequent testing, for example a sub-chronic test, may raise concerns that the more sensitive sex had not been used. In such cases, and only when considerable differences between the sexes are suspected, it may be necessary to conduct another full acute oral toxicity study in the second sex. This is preferable to conducting confirmatory testing in a small group of animals of the second sex as a late satellite to the original test because there is a strong possibility that this would produce results that are difficult to interpret. The impact of conducting a second full test on the overall number of animals used in acute toxicity testing should be small because re-testing is anticipated to be infrequent and the results of the test in one sex, together with data from any subsequent studies, will greatly assist in the selection of starting doses closer to the LD50 in the second test.

14. Healthy young adult animals of commonly used laboratory strains should be employed. Females should be nulliparous and non-pregnant. At the commencement of its dosing, each animal should be between 8 and 12 weeks old and its weight should fall in an interval ± 20 % of the mean initial weight of all previously dosed animals.

Housing and feeding conditions

15. The temperature in the experimental animal room should be 22°C (± 3°C). The relative humidity should be at least 30 % and preferably not exceed 70 % other than during room cleaning. Lighting should be artificial, the sequence being 12 hours light and 12 hours dark. The animals are housed individually. For feeding, conventional rodent laboratory diets may be used with an unlimited supply of drinking water.

Preparation of animals

16. The animals are randomly selected, marked to permit individual identification, and kept in their cages for at least 5 days prior to dosing to allow for acclimatization to the laboratory conditions. As with other sequential test designs, care must be taken to ensure that animals are available in the appropriate size and age range for the entire study.

Preparation of doses

17. When necessary, the test substance is dissolved or suspended in a suitable vehicle. The use of an aqueous solution/suspension/emulsion is recommended wherever possible, followed in order of preference by a solution/suspension/emulsion in oil (e.g. corn oil) and then possibly solution in other vehicles. For vehicles other than water the toxicological characteristics of the

vehicle should be known. Dosing preparations must be prepared shortly prior to administration unless the stability of the preparation over the period during which it will be used is known. Where preparation shortly before administration is not practicable and the stability of the preparation is not known, this will need to be demonstrated analytically.

18. Constant concentration should be used in dosing unless there is clear scientific or regulatory justification for not dosing so. In the event that constant volume was used, information on the actual concentrations used should be provided. In either case, the maximum dose volume for administration must not be exceeded. The maximum volume of liquid that can be administered at one time depends on the size of the test animal. In rodents, the volume should not normally exceed 1 ml/100g of body weight; however, in the case of aqueous solutions, 2 ml/100g body weight can be considered.

Administration of doses

19. The test substance is administered in a single dose by gavage using a stomach tube or a suitable intubation cannula. In the unusual circumstance that a single dose is not possible, the dose may be given in smaller fractions over a period not exceeding 24 hours.

20. Animals should be fasted prior to dosing (e.g., with the rat, food but not water should be withheld overnight; with the mouse, food but not water should be withheld for 3-4 hours). Following the period of fasting, the animals should be weighed and the test substance administered. The fasted body weight of each animal is determined and the dose is calculated according to the body weight. After the substance has been administered, food may be withheld for a further 3-4 hours in rats or 1-2 hours in mice. Where a dose is administered in fractions over a period of time, it may be necessary to provide the animals with food and water depending on the length of the period.

PROCEDURE

Limit test and main test

21. The limit test is primarily used in situations where the experimenter has information indicating that the test material is likely to be nontoxic, i.e., having toxicity below regulatory limit doses. Information about the toxicity of the test material can be gained from knowledge about similar tested compounds or similar tested mixtures or products, taking into consideration the identity and percentage of components known to be of toxicological significance. In those situations where there is little or no information about its toxicity, or in which the test material is expected to be toxic, the main test should be performed.

Limit test

Limit test at 2000 mg/kg

22. Dose one animal at the test dose. If the animal dies, conduct the main test to determine the LD50. If the animal survives, dose four additional animals. The second and third animals can be dosed concurrently and the fourth and fifth sequentially. However, if three animals die, the limit test is terminated and the main test is performed. If an animal unexpectedly dies late in the study, and there are other survivors, it is appropriate to stop dosing and observe all animals to see if other animals will also die during a similar observation period (see paragraph 29 for initial observation period). Late deaths should be counted the same as other deaths. The results are evaluated as follows (O=survival, X=death).

23. The LD50 is less than the test dose (2000 mg/kg) when three or more animals die. If a third animal dies, conduct the main test.

O XO XX
O OX XX
O XX OX
O XX X

24. Test five animals. The LD50 is greater than the test dose (2000 mg/kg) when three or more animals survive.

O OO OO
O OO XO
O OO OX
O OO XX
O XO XO
O XO OO/X
O OX OO/X
O XX OO

<u>Limit Test at 5000 mg/kg</u>

25. Exceptionally, and only when justified by specific regulatory needs, the use of a dose at 5000 mg/kg may be considered (see Appendix D). Recognizing the need to protect animal welfare, testing of animals in class 5 ranges is discouraged and should only be considered when there is a strong likelihood that results of such a test have a direct relevance for protecting human or animal health or the environment.

26. Dose one animal at the test dose. If the animal dies, conduct the main test to determine the LD50. If the animal survives, dose two additional animals. If both animals survive, the LD50 is greater than the limit dose and the test is terminated (i.e. carried to full 14-day observation without dosing of further animals). If one or both animals die, then dose an additional two animals, one at a time. If an animal unexpectedly dies late in the study, and there

are other survivors, it is appropriate to stop dosing and observe all animals to see if other animals will also die during a similar observation period (see paragraph 10 for initial observation period). Late deaths should be counted the same as other deaths. The results are evaluated as follows (O=survival, X=death, and U=Unnecessary).

27. The LD50 is less than the test dose (5000 mg/kg) when three or more animals die.

O XO XX
O OX XX
O XX OX
O XX XU (U can be O or X, the dosing of the 5th animal is not necessary)

28. The LD50 is greater than the test dose (5000 mg/kg) when three or more animals survive.

O OO UU (U can be O or X, the dosing of the 4th and 5th animal is not necessary)
O XO XO
O XO OU (U can be O or X, the dosing of the 5th animal is not necessary)
O OX XO
O OX OU (U can be O or X, the dosing of the 5th animal is not necessary)
O XX OO

Main test

29. Single animals are dosed in sequence usually at 48 h intervals. However, the time intervals between dosing is determined by the onset, duration, and severity of toxic signs. Treatment of an animal at the next dose should be delayed until one is confident of survival of the previously dosed animal. The time interval may be adjusted as appropriate, e.g., in case of inconclusive response. The test is simpler to implement when a single time interval is used for making sequential dosing decisions. Nevertheless, it is not necessary to recalculate dosing or likelihood-ratios if the time interval changes midtest. For selecting the starting dose, all available information, including information on structurally related substances and results of any other toxicity tests on the test material, should be used to approximate the LD50 as well as the slope of the dose-response curve.

30. The first animal is dosed a step below the toxicologist's best estimate of the LD50. If the animal survives, the second animal receives a higher dose. If the first animal dies or appears moribund, the second animal receives a lower dose. The dose progression factor should be chosen to be the antilog of 1/(the estimated slope of the dose-response curve) (a progression of 3.2 corresponds to a slope of 2) and should remain constant throughout testing. Thus, when there is no information on the slope of the substance to be tested, a default dose progression factor of 3.2 is used. Using the default progression factor, doses would be selected from the sequence 1.75, 5.5, 17.5, 55, 175, 550, 2000 (or 1.75, 5.5, 17.5, 55, 175, 550, 1750, 5000 for specific regulatory needs). If no estimate of the substance's lethality is available, dosing should be initiated at 175 mg/kg. In most cases, this dose is sublethal and therefore serves to reduce the

level of pain and suffering. If animal tolerances to the chemical are expected to be highly variable (i.e., slopes are expected to be less than 2.0), consideration should be given to increasing the dose progression factor beyond the default 0.5 on a log dose scale (i.e., 3.2 progression factor) prior to starting the test. Similarly, for test substances known to have very steep slopes, dose progression factors smaller than the default should be chosen. (Appendix B includes a table of dose progressions for whole number slopes ranging from 1 to 8 with starting dose 175 mg/kg.)

31. Dosing continues depending on the fixed-time interval (e.g., 48-hour) outcomes of all the animals up to that time. The testing stops when one of the following stopping criteria first is met:

(a) 3 consecutive animals survive at the upper bound;

(b) 5 reversals occur in any 6 consecutive animals tested;

(c) at least 4 animals have followed the first reversal and the specified likelihood-ratios exceed the critical value. (See paragraph 42 and Appendix C. Calculations are made at each dosing, following the fourth animal after the first reversal.).

For a wide variety of combinations of LD50 and slopes, stopping rule (c) will be satisfied with 4 to 6 animals after the test reversal. In some cases for chemicals with shallow slope dose-response curves, additional animals (up to a total of fifteen tested) may be needed.

32. When the stopping criteria have been attained, the estimated LD50 should be calculated from the animal outcomes at test termination using the method described in paragraphs 38 and 39.

33. Moribund animals killed for humane reasons are considered in the same way as animals that died on test. If an animal unexpectedly dies late in the study and there are other survivors at that dose or above, it is appropriate to stop dosing and observe all animals to see if other animals will also die during a similar observation period. If subsequent survivors also die, AND it appears that all dose levels exceed the LD50 it would be most appropriate to start the study again beginning at least two steps below the lowest dose with deaths (and increasing the observation period) since the technique is most accurate when the starting dose is below the LD50. If subsequent animals survive at or above the dose of the animal that dies, it is not necessary to change the dose progression since the information from the animal that has now died will be included into the calculations as a death at a lower dose than subsequent survivors, pulling the LD50 down.

Observations

34. Animals are observed individually at least once during the first 30 minutes after dosing, periodically during the first 24 hours (with special attention given during the first 4 hours), and

daily thereafter, for a total of 14 days, except where they need to be removed from the study and humanely killed for animal welfare reasons or are found dead. However, the duration of observation should not be fixed rigidly. It should be determined by the toxic reactions and time of onset and length of recovery period, and may thus be extended when considered necessary. The times at which signs of toxicity appear and disappear are important, especially if there is a tendency for toxic signs to be delayed (19). All observations are systematically recorded with individual records being maintained for each animal.

35. Additional observations will be necessary if the animals continue to display signs of toxicity. Observations should include changes in skin and fur, eyes and mucous membranes, and also respiratory, circulatory, autonomic and central nervous systems, and somatomotor activity and behavior pattern. Attention should be directed to observations of tremors, convulsions, salivation, diarrhea, lethargy, sleep and coma. The principles and criteria summarized in the Humane Endpoints Guidance Document (13) should be taken into consideration. Animals found in a moribund condition and animals showing severe pain and enduring signs of severe distress should be humanely killed. When animals are killed for humane reasons or found dead, the time of death should be recorded as precisely as possible.

Body weight

36. Individual weights of animals should be determined shortly before the test substance is administered and at least weekly thereafter. Weight changes should be calculated and recorded. At the end of the test surviving animals are weighed and then humanely killed.

Pathology

37. All animals (including those which die during the test or are removed from the study for animal welfare reasons) should be subjected to gross necropsy. All gross pathological changes should be recorded for each animal. Microscopic examination of organs showing evidence of gross pathology in animals surviving 24 or more hours after the initial dosing may also be considered because it may yield useful information.

DATA AND REPORTING

Data

38. Individual animal data should be provided. Additionally, all data should be summarized in tabular form, showing for each test dose the number of animals used, the number of animals displaying signs of toxicity (19), the number of animals found dead during the test or killed for humane reasons, time of death of individual animals, a description and the time course of toxic effects and reversibility, and necropsy findings. A rationale for the starting dose and the dose progression and any data used to support this choice should be provided.

Calculation of LD50 for the main test

39. The LD50 is calculated using the maximum likelihood method (17)(18), except in the exceptional cases described in paragraph 40. The following statistical details may be helpful in implementing the maximum likelihood calculations suggested (with an assumed *sigma*). All deaths, whether immediate or delayed or humane kills, are incorporated for the purpose of the maximum likelihood analysis. Following Dixon (4), the likelihood function is written as follows:

$L = L_1 L_2 \ldots L_n$,

where

L is the likelihood of the experimental outcome, given *mu* and *sigma*, and n the total number of animals tested.

$L_i = 1 - F(Z_i)$ if the ith animal survived, or
$L_i = F(Z_i)$ if the ith animal died,

where

F = cumulative standard normal distribution,
$Z_i = [\log(d_i) - mu\ ] / sigma$
d_i = dose given to the ith animal, and
sigma = standard deviation in log units of dose (which is not the log standard deviation.

An estimate of the log of the true LD50 is given by the value of *mu* that maximizes the likelihood L (see paragraph 41).

An estimate of *sigma* of 0.5 is used unless a better generic or case-specific value is available.

40. Under some circumstances, statistical computation will not be possible or will likely give erroneous results. Special means to determine/report an estimated LD50 are available for these circumstances as follows:

(a) If testing stopped based on criterion (a) in paragraph 31 (i.e., a boundary dose was tested repeatedly), or if the upper bound dose ended testing, then the LD50 is reported to be above the upper bound.

(b) If all the dead animals have higher doses than all the live animals (or if all live animals have higher doses than all the dead animals, although this is practically unlikely), then the LD50 is between the doses for the live and the dead animals. These observations give no further information on the exact value of the LD50. Still, a maximum likelihood LD50 estimate can be made provided there is a value for *sigma*.

Stopping criterion (b) in paragraph 31 describes one such circumstance.

(c) If the live and dead animals have only one dose in common and all the other dead animals have higher doses and all the other live animals lower doses, or vice versa, then the LD50 equals their common dose. If a closely related substance is tested, testing should proceed with a smaller dose progression.

If none of the above situations occurs, then the LD50 is calculated using the maximum likelihood method.

41. Maximum likelihood calculation can be performed using either SAS (17) (e.g., PROC NLIN) or BMDP (18) (e.g., program AR) computer program packages as described in Appendix 1D in Reference 3. Other computer programs may also be used. Typical instructions for these packages are given in appendices to the ASTM Standard E 1163-87 (6). (The *sigma* used in the BASIC program in (6) will need to be edited to reflect the parameters of the Up-and-Down Procedure.) The program's output is an estimate of log(LD50) and its standard error.

42. The likelihood-ratio stopping rule (c) in paragraph 31 is based on three measures of test progress, that are of the form of the likelihood in paragraph 39, with different values for *mu*. Comparisons are made after each animal tested after the sixth that does not already satisfy criterion (a) or (b) of paragraph 31. The equations for the likelihood-ratio criteria are provided in Appendix C. These comparisons are most readily performed in an automated manner and can be executed repeatedly, for instance, by a spreadsheet routine such as that also provided in Appendix C. If the criterion is met, testing stops and the LD50 can be calculated by the maximum likelihood method.

Computation of confidence interval

(43) Following the main test and estimated LD50 calculation, it may be possible to compute interval estimates for the LD50. Any of these confidence intervals provides valuable information on the reliability and utility of the main test that was conducted. A wide confidence interval indicates that there is more uncertainty associated with the estimated LD50. In this case, the reliability of the estimated LD50 is low and the usefulness of the estimated LD50 may be marginal. A narrow interval indicates that there is relatively little uncertainty associated with the estimated LD50. In this case, the reliability of the estimated LD50 is high and the usefulness of the estimated LD50 is good. This means that if the main test were to be repeated, the new estimated LD50 is expected to be close to the original estimated LD50 and both of these estimates are expected to be close to the true LD50.

Depending on the outcome of the main test, one of two different types of interval estimates of the true LD50 is calculated:

(a) When at least three different doses have been tested and the middle dose has at least one animal that survived and one animal that died, a profile-likelihood-based

computational procedure is used to obtain a confidence interval that is expected to contain the true LD50 95% of the time. However, because small numbers of animals are expected to be used, the actual level of confidence is generally not exact (20). The random stopping rule improves the ability of the test overall to respond to varying underlying conditions, but also causes the reported level of confidence and the actual level of confidence to differ somewhat (21).

(b) If all animals survive at or below a given dose level and all animals die when dosed at the next higher dose level, an interval is calculated that has as its lower limit the highest dose tested where all the animals survive and has as its upper limit the dose level where all the animals died. This interval is labeled as "approximate." The exact confidence level associated with this interval cannot be specifically determined. However, because this type of response would only occur when the dose-response is steep, in most cases, the true LD50 is expected to be contained within the calculated interval or be very close to it. This interval will be relatively narrow and sufficiently accurate for most practical use.

In some instances, confidence intervals are reported as infinite, through including either zero at the lower end or infinity at the upper end, or both. Such intervals may occur, for example, when the response profile is relatively flat or relatively uncertain.

Implementing this set of procedures requires specialized computation which is either by use of a dedicated program to be available through the Environmental Protection Agency (EPA) or OECD or developed following technical details available from the EPA or OECD. Achieved coverage of these intervals and properties of the dedicated program are described in a report (22) also available through the EPA. Appendix E provides information on choice of dose progression and initial dose level for the UDP and describes test performance under a variety of circumstances.

Report

44. The test report must include the following information:

Test substance:

- physical nature, purity and physicochemical properties (including isomerization);
- identification data.

Vehicle (if appropriate):

- justification for choice of vehicle, if other than water.

Test animals:

- species/strain used;
- microbiological status of the animals, when known;
- number, age and sex of animals;
- rationale for use of males instead of females;
- source, housing conditions, diet, etc.;
- individual weights of animals at the start of the test, at day 7, and at day 14.

Test conditions:

- rationale for initial dose level selection, dose progression factor and for follow-up dose levels;
- details of test substance formulation;
- details of the administration of the test substance;
- details of food and water quality (including diet type/source, water source).

Results:

- body weight/body weight changes;
- tabulation of response data by sex (if both sexes are used) and dose level for each animal (i.e., animals showing signs of toxicity including nature, severity, duration of effects, and mortality);
- time course of onset of signs of toxicity and whether these were reversible for each animal;
- necropsy findings and any histopathological findings for each animal, if available;
- LD50 data;
- statistical treatment of results (description of computer routine used and spreadsheet tabulation of calculations)

Discussion and interpretation of results.

Conclusions.

LITERATURE

(1) Dixon, W.J. and A.M. Mood. (1948). A Method for Obtaining and Analyzing Sensitivity Data. J. Amer. Statist. Assoc., 43, 109-126.

(2) Dixon, W.J. (1965). The Up-and-Down Method for Small Samples. J. Amer. Statist. Assoc. 60, 967-978.

(3) Dixon, W.J. (1991). Staircase Bioassay: The Up-and-Down Method. Neurosci. Biobehav. Rev., 15, 47-50.

(4) Dixon, W.J. (1991). Design and Analysis of Quantal Dose-Response Experiments (with Emphasis on Staircase Designs). Dixon Statistical Associates, Los Angeles CA, USA.

(5) Bruce, R.D. (1985). An Up-and-Down Procedure for Acute Toxicity Testing. Fundam. Appl. Tox., 5, 151-157.

(6) ASTM (1987). E 1163-87, Standard Test Method for Estimating Acute Oral Toxicity in Rats. American Society for Testing and Materials, Philadelphia Pa, USA.

(7) Lipnick, R.L., J.A. Cotruvo, R.N. Hill, R.D. Bruce, K.A. Stitzel, A.P. Walker, I. Chu, M. Goddard, L. Segal, J.A. Springer, and R.C. Myers. (1995). Comparison of the Up-and-Down, Conventional LD50 and Fixed Dose Acute Toxicity Procedures. Fd. Chem. Toxicol., 33, 223-231.

(8) Choi, S.C. (1990). Interval Estimation of the LD50 Based on An Up-and-Down Experiment. Biometrics 46, 485-492.

(9) Vågerö, M. and R. Sundberg. (1999). The Distribution of The Maximum Likelihood Estimator in Up-and-Down Experiments for Quantal Dose-Response Data. J. Biopharmaceut. Statist. 9(3), 499-519.

(10) Hsi, B.P. (1969). The Multiple Sample Up-and-Down Method in Bioassay. J. Amer. Statist. Assoc. 64, 147-162.

(11) Noordwijk, A.J. van and J. van Noordwijk. (1988). An Accurate Method for Estimating An Approximate Lethal Dose with Few Animals, Tested with A Monte Carlo Procedure. Arch. Toxicol. 61, 333-343.

(12) OECD (2000). Guidance Document on Acute Oral Toxicity. Environmental Health and Safety Monograph Series on Testing and Assessment No. 24.

(13) OECD. (2000). Revised Draft Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation.

(14) Report of the International Workshop on *In Vitro* Methods for Assessing Acute Systemic Toxicity. NIH Publication No. 01-4499. Research Triangle Park, NC: National Institute of Environmental Health Sciences, August 2001.

(15) Guidance Document on Using *In Vitro* Data to Estimate *In Vivo* Starting Doses for Acute Toxicity. NIH Publication No. 01-4500. Research Triangle Park, NC: National Institute of Environmental Health Sciences, August 2001.

(16) Spielmann, H.E., M.Genschow, M. Leibsch, and W. Halle. (1999). Determination of the

Starting Dose for Acute Oral Toxicity (LD50) Testing in the Up-and-Down Procedure (UDP) from Cytotoxicity Data ATLA 27: 957-966.

(17) SAS Institute Inc. (1990). SAS/STAT® User's Guide. Version 6, Fourth Ed. or later. Cary, NC, USA.

(18) BMDP Statistics Software, Inc. (1990). BMDP Statistical Software Manual. W.J. Dixon, Chief Ed. 1990 rev. or later. University of California Press, Berkeley, CA, USA.

(19) Chan, P.K. and A.W. Hayes. (1994). Chap. 16. Acute Toxicity and Eye Irritancy. *Principles and Methods of Toxicology*. Third Edition. A.W. Hayes, Editor. Raven Press, Ltd., New York, USA.

(20) Rosenberger, W.F., Flournoy, N. and Durham, S.D. (1997). Asymptotic Normality of Maximum Likelihood Estimators from Multiparameter Response-driven Designs. Journal of Statistical Planning and Inference 60, 69-76.

(21) Jennison, C. and B.W. Turnbull. (2000). *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall/CRC: Boca Raton, FL.

(22) Westat. (2001). Simulation Results for the AOT425StatPgm Program. Report prepared for U.S. E.P.A. under Contract 68-W7-0025, Task Order 5-03.

APPENDIX A

DEFINITIONS

Acute oral toxicity is the adverse effects occurring within a short time of oral administration of a single dose of a substance or multiple doses given within 24 hours.

Confidence interval is an interval estimate, a range of values, intended to include the true LD50 with a specified degree of confidence.

Delayed death means that an animal does not die or appear moribund within 48 hours but dies later during the 14-day observation period.

Dose is the amount of test substance administered. Dose is expressed as weight (g, mg) or as weight of test substance per unit weight of test animal (e.g., mg/kg).

Dose progression factor, sometimes termed a dose spacing factor, refers to the multiple by which a dose is increased (i.e., the dose progression) when an animal survives or the divisor by which it is decreased when an animal dies. The dose progression factor is recommended to be the antilog of 1/(the estimated slope of the dose-response curve). The default dose progression factor is recommended to be 3.2 = antilog 0.5 = antilog (½).

LD50 (median lethal dose), oral, is a statistically derived single dose of a substance that can be expected to cause death in 50 per cent of animals when administered by the oral route. The LD50 value is expressed in terms of weight of test substance per unit weight of test animal (mg/kg).

Limit dose refers to a dose at an upper limitation on testing (2000-5000 mg/kg).

Moribund status of an animal refers to being in a state of dying or inability to survive, even if treated.

Nominal sample size refers to the total number of tested animals**,** reduced by one less than the number of like responses at the beginning of the series, or by the number of tested animals up to but not including the pair that creates the first reversal. For example, for a series where X and O indicate opposite animal outcomes (for instance, X could be dies within 48 hours and O survives) in a pattern as follows: OOOXXOXO, we have the total number of tested animals (or sample size in the conventional sense) as 8 and the nominal sample size as 6. This particular example shows 4 animals following a reversal. It is important to note whether a count in a particular part of the guideline refers to the nominal sample size or to the total number tested. For example, the maximum actual number tested is 15. When testing is stopped based on that basis, the nominal sample size will be less than or equal to 15. Members of the nominal sample start with the (r-1)st animal (the animal before the second in the reversal pair) (see reversal below).

Probity is an abbreviation for the term "probability integral transformation" and a probity dose-response model permits a standard normal distribution of expected responses (i.e., one centered to its mean and scaled to its standard deviation, *sigma*) to doses (typically in a logarithmic scale) to be analyzed as if it were a straight line with slope the reciprocal of *sigma*. A standard normal lethality distribution is symmetric; hence, its mean is also its true LD50 or median response.

Reversal is a situation where nonresponse is observed at some dose, and a response is observed at the next dose tested, or vice versa (i.e., response followed by nonresponse). Thus, a reversal is created by a pair of responses. The first such pair occurs at animals numbered r-1 and r.

Sigma is the standard deviation of a log normal curve describing the range of tolerances of test subjects to the chemical (where a subject is expected capable of responding if the chemical dose exceeds the subject's tolerance). The estimated *sigma* provides an estimate of the variation among test animals in response to a full range of doses. See slope and probity.

Slope (of the dose-response curve) is a value related to the angle at which the dose response curve rises from the dose axis. In the case of probity analysis, when responses are analyzed on a probity scale against dose on a log scale this curve will be a straight line and the slope is the reciprocal of *sigma,* the standard deviation of the underlying test subject tolerances, which are assumed to be normally distributed. See probity and *sigma.*

Stopping rule is used in this guideline synonymously with 1) a specific stopping criterion and 2) the collection of all criteria determining when a testing sequence terminates. In particular, for the main test, stopping rule is used in paragraph 5 as a shorthand for the criterion that relies on comparison of ratios to a critical value.

APPENDIX B

DOSING PROCEDURE

Dose Sequence for Main Test

1. Up-and-Down Dosing Procedure. For each run, animals are dosed, one at a time, usually at 48-hour intervals. The first animal receives a dose a step below the level of the best estimate of the LD50. This selection reflects an adjustment for a tendency to bias **away from the LD50 in the direction of** the initial starting dose in the final estimate (see paragraph 5 of the test guideline). The overall pattern of outcomes is expected to stabilize as dosing is adjusted for each subsequent animal. Paragraph 3 below provides further guidance for choice of dose spacing factor).

2. Default Dose Progression. Once the starting dose and dose spacing are decided, the toxicologist should list all possible doses including the upper bound (usually 2000 or 5000 mg/kg). Doses that are close to the upper bound should be removed from the progression. The stepped nature of the Up-and-Down Procedure design provides for the first few doses to function as a self-adjusting sequence. Because of the tendency for positive bias, in the event that nothing is known about the substance, a starting dose of 175 mg/kg is recommended. If the default procedure is to be used for the main test, dosing will be initiated at 175 mg/kg and doses will be spaced by a factor of 0.5 on a log dose scale. The doses to be used include 1.75, 5.5, 17.5, 55, 175, 550, 2000 or, for specific regulatory needs, 1.75, 5.5, 17.5, 55, 175, 550, 1750, 5000. For certain highly toxic substances, the dosing sequence may need to be extended to lower values.

3. In the event a dose progression factor other than the default is deemed suitable, Table B.1 provides dose progressions for whole number multiples of slope, from 1 to 8.

Table B.1. Dose Progressions for Up-and-Down Procedure
Choose a Slope and Read Down the Column
All doses in mg/kg bw

Slope =	1	2	3	4	5	6	7	8
	0.175*	0.175*	0.175*	0.175*	0.175*	0.175*	0.175*	0.175*
							0.243	0.233
					0.28	0.26		
				0.31			0.34	0.31
			0.38			0.38		
								0.41
					0.44		0.47	
		0.55		0.55		0.55		0.55
					0.70		0.65	
								0.74
			0.81			0.81		
				0.98			0.91	0.98
					1.10	1.19		
							1.26	1.31
	1.75	1.75	1.75	1.75	1.75	1.75	1.75	1.75
							2.43	2.33
					2.8	2.6		
				3.1			3.4	3.1
			3.8			3.8		
					4.4			4.1
							4.7	
		5.5		5.5		5.5		5.5
					7.0		6.5	
								7.4
			8.1			8.1		
				9.8			9.1	9.8
					11.0	11.9		
							12.6	13.1
	17.5	17.5	17.5	17.5	17.5	17.5	17.5	17.5
							24.3	23.3
					28	26		
				31			34	31

Table B.1. (continued)

Slope =	1	2	3	4	5	6	7	8
			38			38		
					44			41
							47	
		55		55		55		55
							65	
					70			74
			81			81		
				98			91	98
					110	119		
							126	131
	175	175	175	175	175	175	175	175
							243	233
					280	260		
				310			340	310
			380			380		
					440			410
							470	
		550		550		550		550
							650	
					700			740
			810			810		
				980			910	980
					1100	1190		
							1260	1310
	1750	1750	1750	1750	1750	1750	1750	1750
							2430	2330
					2800	2600		
				3100				3100
						3800	3400	
								4100
	5000	5000	5000	5000	5000	5000	5000	5000

* If lower doses are needed, continue progressions to a lower dose

APPENDIX C

COMPUTATIONS FOR THE LIKELIHOOD-RATIO STOPPING RULE

As described in Guideline paragraph 31, the main test may be completed on the basis of the first of three stopping criteria to occur. In any case, even if none of the stopping criteria are satisfied, dosing would stop when 15 animals are dosed. Tables C.1 - C.4 illustrate examples where testing has started with no information, so the recommended default starting value, 175 mg/kg, and the recommended default dose progression factor, 3.2 or one half log, have been used.

Table C.1 shows how the main test would stop if 3 animals have survived at the limit dose of 2000 mg/kg; Table C.2 shows a similar situation when the limit dose of 5000 mg/kg is used. (These illustrate situations where a Limit Test was not thought appropriate *a priori*.) Table C.3 shows how a particular sequence of 5 reversals in 6 tested animals could occur and allow test completion. Finally, Table C.4 illustrates a situation several animals into a test, where neither criterion (a) nor criterion (b) has been met, a reversal of response has occurred followed by 4 tested animals, and, consequently, criterion (c) must be evaluated as well.

Criterion (c) calls for a likelihood-ratio stopping rule to be evaluated after testing each animal, starting with the fourth tested following the reversal. Three "measures of test progress" are calculated. Technically, these measures of progress are likelihoods, as recommended for the maximum-likelihood estimation of the LD50. The procedure is closely related to calculation of a CI by a likelihood-based procedure.

The basis of the procedure is that when enough data have been collected, a point estimate of the LD50 should be more strongly supported than values above and below the point estimate, where statistical support is quantified using likelihood. Therefore three likelihood values are calculated: a likelihood at an LD50 point estimate (called the rough estimate or dose-averaging estimate in the example), a likelihood at a value below the point estimate, and a likelihood at a value above the point estimate. Specifically, the low value is taken to be the point estimate divided by 2.5 and the high value is taken to be the point estimate multiplied by 2.5.

The likelihood values are compared by calculating ratios of likelihoods, and then determining whether these likelihood-ratios (LR) exceed a critical value. Testing stops when the ratio of the likelihood for the point estimate exceeds each of the other likelihoods by a factor of 2.5, which is taken to indicate relatively strong statistical support for the point estimate. Therefore two likelihood-ratios (LRs) are calculated, a ratio of likelihoods for the point estimate and the point estimate divided by 2.5, and a ratio for the point estimate and the estimate times 2.5.

The calculations are easily performed in any spreadsheet with normal probability functions. The calculations are illustrated in Table C.4, which is structured to promote spreadsheet implementation. The computation steps are illustrated using an example where the upper limit dose is 5000 mg/kg, but the computational steps are carried out in the same fashion when the

upper boundary dose is 2000 mg/kg. Empty spreadsheets preprogrammed with the necessary formulas are available for direct downloading on the OECD and EPA web sites.

Hypothetical example using an upper limit dose of 5000 mg/kg (Table C.4)
In the hypothetical example utilizing an upper boundary dose of 5000 mg/kg, the LR stopping criterion was met after nine animals had been tested. The first "reversal" occurred with the 3rd animal tested. The LR stopping criterion is checked when four animals have been tested following the reversal. In this example, the fourth animal tested following the reversal is the seventh animal actually tested. Therefore, for this example, the spreadsheet calculations are only needed after the seventh animal had been tested and the data could be entered at that time. Subsequently, the LR stopping criterion would have been checked after testing the seventh animal, the eighth animal, and the ninth. The LR stopping criterion is first satisfied after the ninth animal is tested in this example.

A. Enter the dose-response information animal by animal.

Column 1. Steps are numbered 1-15. No more than 15 animals may be tested.
Column 2. Place an I in this column as each animal is tested.
Column 3. Enter the dose received by the ith animal.
Column 4. Indicate whether the animal responded (shown by an X) or did not respond (shown by an O).

B. The nominal and actual sample sizes.

The nominal sample consists of the two animals that represent the first reversal (here the second and third animals), plus all animals tested subsequently. Here, Column 5 indicates whether or not a given animal is included in the nominal sample.

- The nominal sample size (nominal *n*) appears in Row 16. This is the number of animals in the nominal sample. In the example, nominal *n* is 8.
- The actual number tested appears in Row 17.

C. Rough estimate of the LD50.

The geometric mean of doses for the animals in the current nominal sample is used as a rough estimate of the LD50 from which to gauge progress. In the table, this is called the "dose-averaging estimator." It is updated with each animal tested. This average is restricted to the nominal sample in order to allow for a poor choice of initial test dose, which could generate either an initial string of responses or an initial string of nonresponses. (However, the results for all animals are used in the likelihood calculations for final LD50 calculation below.) Recall that the geometric mean of *n* numbers is the product of the *n* numbers, raised to a power of 1/*n*.

- The dose-averaging estimate appears in Row 18 (e.g., $(175 * 550 * ... * 1750)^{1/8} = 1292.78$).

A Row 19 shows the logarithm (base 10) of the value in Row 18 (e.g., $\log_{10} 1292.8 = 3.112$).

D. Likelihood for the rough LD50 estimate.

"Likelihood" is a statistical measure of how strongly the data support an estimate of the LD50 or other parameter. Ratios of likelihood values can be used to compare how well the data support different estimates of the LD50.

In Column 8 calculate the likelihood for Step C's rough LD50 estimate. The likelihood (Row 21) is the product of likelihood contributions for individual animals (see Guideline paragraph 39). The likelihood contribution for the ith animal is denoted L_i.

Column 7. Enter the estimate of the probability of response at dose d_i, denoted P_i. P_i is calculated from a dose-response curve. Note that the parameters of a probity dose-response curve are the slope and the LD50, so values are needed for each of those parameters. For the LD50 the dose-averaging estimate from Row 18 is used. For the slope in this example the default value of 2 is used. The following steps may be used to calculate the response probability P_i.

1. Calculate the base-10 log of dose d_i (Column 6).

2. For each animal calculate the z-score, denoted Z_i (not shown in the table), using the formulae
   $sigma = 1 / slope$,
   $Z_i = (\log_{10}(d_i) - \log_{10}(LD50)) / sigma$

   For example, for the first animal (Row 1),
   $sigma = 1 / 2$
   $Z_1 = (2.243 - 3.112) / 0.500 = -1.738$

3. For the ith dose the estimated response probability is

   $P_i = F(Z_i)$

where F denotes the cumulative distribution function for the standard normal distribution (i.e., the normal distribution with mean 0 and variance 1).

For example (Row 1),

$P_1 = F(-1.738) = 0.0412$

The function F (or something very close) is ordinarily what is given for the normal distribution in statistical tables, but the function is also widely available as a spreadsheet function. It is

available under different names, for example the @NORMAL function of Lotus 1-2-3 (1) and the @NORMDIST function in Excel (2). To confirm that you have used correctly the function available in your software, you may wish to verify familiar values such as $F(1.96)$. 0.975 or $F(1.64)$. 0.95.

Column 8. Calculate the natural log of the likelihood contribution ($\ln(L_i)$). L_i is simply the probability of the response that actually was observed for the ith animal:

responding animals: $\ln(L_i) = \ln(P_i)$
non-responding animals: $\ln(L_i) = \ln(1 - P_i)$

Note that here the natural logarithm (ln) is used, whereas elsewhere the base-10 (common) logarithm was used. These choices are what are ordinarily expected in a given context.

The steps above are performed for each animal. Finally:

Row 20: Sum the log-likelihood contributions in Column 8.
Row 21: Calculate the likelihood by applying the exp function applied to the log-likelihood value in Row 20 (e.g., $\exp(-3.389) = e^{-3.389} = 0.0337$).

E. <u>Calculate likelihoods for two dose values above and below the rough estimate.</u>

If the data permit a precise estimate, then one expects the likelihood should be high if the estimate is a reasonable estimate of the LD50, relative to likelihoods for values distant from this estimate. Compare the likelihood for the dose-averaging estimate (1292.8, Row 18) to values differing by a factor of 2.5 from that value (i.e., to 1292.8*2.5 and 1292.8/2.5). The calculations (displayed in Columns 9-12) are carried out in a fashion similar to those described above, except that the values 517.1 (=1292.8/2.5) and 3232.0 (=1292.8*2.5) have been used for the LD50, instead of 1292.8. The likelihoods and log-likelihoods are displayed in Rows 20-21.

F. <u>Calculate likelihood-ratios.</u>

The three likelihood values (Row 21) are used to calculate two likelihood-ratios (Row 22). A likelihood-ratio is used to compare the statistical support for the estimate of 1292.8 to the support for each of the other values, 517.1 and 3232.0. The two likelihood-ratios are therefore:

LR1 = [likelihood of 1292.8] / [likelihood of 517.1]
= 0.0337 / 0.0080
= 4.21

and

LR2 = [likelihood of 1292.8] / [likelihood of 3232.0]
= 0.0337 / 0.0098
= 3.44

G. Determine if the likelihood-ratios exceed the critical value.

High likelihood-ratios are taken to indicate relatively high support for the point estimate of the LD50. Both of the likelihood-ratios calculated in Step F (4.21 and 3.44) exceed the critical likelihood-ratio, which is 2.5. Therefore the LR stopping criterion is satisfied and testing stops. This is indicated by a TRUE in Row 24 and a note at the top of the example spreadsheet that the LR criterion is met.

LITERATURE

(1) Lotus Development Corporation. (1999). Lotus® 1-2-3. Version 9.5, Millennium Edition. Cambridge, MA, USA.

(2) Microsoft Corporation. (1985-1997). Microsoft® Excel. Version 5.0 or later. Seattle, WA, USA.

Table C.1. Example of stopping criterion (a) using 2000 mg/kg.

Stop after animal #5 because 3 animals survive at limit of 2000 mg/kg (#3-#5).

1	2	3	4	5	6	7	8	9	10	11	12
Step	(I)nclude; (E)xclude	Dose	(X)response (O)non-resp. OK	Included in nominal *n*	log10 Dose	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)
1	I	175	O	no	2.2430	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
2	I	550	O	no	2.7404	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
3	I	2000	O	no	3.3010	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
4	I	2000	O	no	3.3010	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
5	I	2000	O	no	3.3010	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
6	E				-						
7	E				-						
8	E				-	-	-	-	-	-	-
9	E				-	-	-	-	-	-	-
10	E				-	-	-	-	-	-	-
11	E				-	-	-	-	-	-	-
12	E				-	-	-	-	-	-	-
13	E							-	-	-	-
14	E							-	-	-	-
15	E							-	-	-	-
Nominal Sample size =				0							
Actual number tested =				5							

Ignore all calculation cells. No reversal in direction of response.

Maximum Likelihood Calculations cannot be completed. LD50 is greater than 2000 mg/kg.

Calculated maximum likelihood estimate of LD50 = none

Table C.2. Example of stopping criterion (a) using 5000 mg/kg.

Stop after animal #6 because 3 animals survive at limit of 5000 mg/kg (#4-#6).

1	2	3	4	5	6	7	8	9	10	11	12
Step	(I)nclude; (E)xclude	Dose	(X)response (O)non-resp. OK	Included in nominal *n*	log10 Dose	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)
1	I	175	O	no	2.2430	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
2	I	550	O	no	2.7404	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
3	I	1750	O	no	3.2430	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
4	I	5000	O	no	3.6990	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
5	I	5000	O	no	3.6990	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
6	I	5000	O	no	3.6990	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
7	E				-						
8	E				-						
9	E				-	-	-	-	-	-	-
10	E				-	-	-	-	-	-	-
11	E				-	-	-	-	-	-	-
12	E				-				-	-	-
13	E				-				-	-	-
14	E				-				-	-	-
15	E				-	-	-	-	-	-	-
Nominal Sample size =				0							
Actual number tested =				6							

Ignore all calculation cells. No reversal in direction of response.

Maximum Likelihood Calculations cannot be completed. LD50 is greater than 5000 mg/kg.

Calculated maximum likelihood estimate of LD50 = none

Table C.3. Example of stopping criterion (b).

Stop after animal #7 because 5 reversals in 6 consecutive animals tested (#2-#7).

1	2	3	4	5	6	7	8	9	10	11	12
Step	(I)nclude; (E)xclude	Dose	(X)response (O)non-resp. OK	Included in nominal *n*	log10 Dose	LD50 = Prob. of response	31.0 likelihood contribn. (ln *Li*)	LD50 = Prob. of response	12.4 likelihood contribn. (ln *Li*)	LD50 = Prob. of response	77.6 likelihood contribn. (ln *Li*)
1	I	175	X	no	2.2430	0.9335	-0.0688	0.9892	-0.0108	0.7602	-0.2742
2	I	55	X	yes	1.7404	0.6905	-0.3703	0.9020	-0.1031	0.3826	-0.9607
3	I	17.5	O	yes	1.2430	0.3095	-0.3703	0.6174	-0.9607	0.0980	-0.1031
4	I	55	X	yes	1.7404	0.6905	-0.3703	0.9020	-0.1031	0.3826	-0.9607
5	I	17.5	O	yes	1.2430	0.3095	-0.3703	0.6174	-0.9607	0.0980	-0.1031
6	I	55	X	yes	1.7404	0.6905	-0.3703	0.9020	-0.1031	0.3826	-0.9607
7	I	17.5	O	yes	1.2430	0.3095	-0.3703	0.6174	-0.9607	0.0980	-0.1031
8	E				-	-	-	-	-	-	-
9	E				-	-	-	-	-	-	-
10	E				-	-	-	-	-	-	-
11	E				-	-	-	-	-	-	-
12	E				-	-	-	-	-	-	-
13	E				-	-	-	-	-	-	-
14	E				-	-	-	-	-	-	-
15	E				-	-	-	-	-	-	-
Nominal Sample size =				6							
Actual number tested =				7							
Dose-averaging estimator				31.02							
log10 =				1.492							
log-likelihood sums:							-2.2906		-3.2021		-3.4655
likelihoods:							0.1012		0.0407		0.0313
likelihood ratios:									2.4880		3.2378
Individual ratios exceed critical value?				critical=	2.5				FALSE		TRUE
Both ratios exceed critical value?									FALSE		
Calculated maximum likelihood estimate of LD50 =					29.6						

Automated calculation; not relevant to this case.

Final estimate obtained from Maximum Likelihood Calculations

Table C.4. Example of stopping criterion (c).

Stop when LR criterion is first met, here at animal #9.
Check LR criterion starting at animal #6.

Assumed slope	2	sigma =	0.5

Parameters of convergence criterion	
critical LR	2.5
factor of LD50	2.5

Result: The LR criterion is met

1	2	3	4	5	6		7	8	9	10	11	12
Step	(I)nclude; (E)xclude	Dose	(X)response (O)non-resp. OK	Included in nominal *n*	log10 Dose	Contrib.to DAE	LD50 = Prob. of response	1292.8 likelihood contribn. (ln *Li*)	LD50 = Prob. of response	517.1 likelihood contribn. (ln *Li*)	LD50 = Prob. of response	3232.0 likelihood contribn. (ln *Li*)
1	I	175	O	no	2.2430	0.0000	0.0412	-0.0421	0.1733	-0.1903	0.0057	-0.0057
2	I	550	O	yes	2.7404	2.7404	0.2289	-0.2600	0.5214	-0.7368	0.0620	-0.0640
3	I	1750	X	yes	3.2430	3.2430	0.6037	-0.5046	0.8552	-0.1564	0.2971	-1.2138
4	I	550	O	yes	2.7404	2.7404	0.2289	-0.2600	0.5214	-0.7368	0.0620	-0.0640
5	I	1750	X	yes	3.2430	3.2430	0.6037	-0.5046	0.8552	-0.1564	0.2971	-1.2138
6	I	550	O	yes	2.7404	2.7404	0.2289	-0.2600	0.5214	-0.7368	0.0620	-0.0640
7	I	1750	O	yes	3.2430	3.2430	0.6037	-0.9257	0.8552	-1.9323	0.2971	-0.3525
8	I	5000	X	yes	3.6990	3.6990	0.8800	-0.1279	0.9756	-0.0247	0.6477	-0.4344
9	I	1750	X	yes	3.2430	3.2430	0.6037	-0.5046	0.8552	-0.1564	0.2971	-1.2138
10	E				-	0.0000	-	-	-	-	-	-
11	E				-	0.0000	-	-	-	-	-	-
12	E				-	0.0000	-	-	-	-	-	-
13	E				-	0.0000	-	-	-	-	-	-
14	E				-	0.0000	-	-	-	-	-	-
15	E				-	0.0000	-	-	-	-	-	-
Nominal Sample size =				8								
Actual number tested =				9								
Dose-averaging estimator				1292.78								
log10 =				3.112								
log-likelihood sums:								-3.3894		-4.8270		-4.6260
likelihoods:								0.0337		0.0080		0.0098
likelihood ratios:										4.2104		3.4436
Individual ratios exceed critical value?				critical=	2.5					TRUE		TRUE
Both ratios exceed critical value?										TRUE		
Calculated maximum likelihood estimate of LD50 =						1329.6						

Final estimate obtained from Maximum Likelihood Calculations

APPENDIX D

CRITERIA FOR CLASSIFICATION OF TEST SUBSTANCES WITH EXPECTED LD50 VALUES EXCEEDING 2000 MG/KG WITHOUT THE NEED FOR TESTING

Test substances could be classified in the hazard classification defined by:
2000 mg/kg<LD50<5000 mg/kg (Class 5 in the Globally Harmonized System (GHS)) in the following cases:

a) if reliable evidence is already available that indicates that LD50 to be in the range of class 5 values; or other animal studies or toxic effects in humans indicate a concern for human health of an acute nature.

b) through extrapolation, estimation or measurement of data if assignment to a more hazardous class is not warranted, and
- reliable information is available indicating significant toxic effects in humans, or
- any mortality is observed when tested up to class 4 values by the oral route, or
- where expert judgement confirms significant clinical signs of toxicity, when tested up to class 4 values, except for diarrhea, piloerection or an ungroomed appearance, or
- where expert judgement confirms reliable information indicating the potential for significant acute effect from the other animal studies.

APPENDIX E

PERFORMANCE OF THE UP-AND-DOWN PROCEDURE

This appendix addresses choice of dose progression and initial dose level for the UDP and describes the performance of the test under a variety of circumstances. A companion document titled "Toxicology Summary: Performance of the Up-and-Down Procedure" provides assistance to the user in interpretation of the test results and is available on the ICCVAM web site at http://iccvam.niehs.nih.gov/methods/udpdocs/udprpt/udp_ciprop.htm. The statistical methods applied will depend upon the case into which the test response patterns fall (see Table E.1).

1. Adjusting the Dose Progression and Initial Dose.

For optimum performance of the UDP, the dose progression used should be based on an accurate estimate of *sigma*. The following two cases describe the outcome when an accurate estimate of *sigma* is not available. In addition, to account conservatively for any bias in the LD50 estimate, it is essential that dosing be initiated below the actual LD50.

(i) Assumed *sigma* << true *sigma*:

When the assumed *sigma* (i.e., the *sigma* on which the dose progression is based) is much smaller than the true *sigma* of the actual test population, the estimated LD50 may be "biased" in the direction of starting dose. For example, if the starting dose is less than the true LD50 of the test population, the estimated LD50 will generally be below the true LD50. Also, if the starting dose is greater than the true LD50 of the test population, the estimated LD50 will tend to be greater than the true LD50. To minimize the chance of overestimating the LD50 due to this bias, the UDP guideline recommends a choice of starting dose just below the assumed LD50.

(ii) Assumed *sigma* >> true *sigma*:

If the assumed *sigma* on which the dose progression is based is much larger than the true *sigma* of the test population, the median estimated LD50 can be much larger or much smaller than the true LD50 depending on the starting dose. In this case, the LD50 can be estimated only within a range. (This is Case 3 described below.)

2. Confidence Interval.

Coverage of the confidence interval is the probability that a calculated confidence interval encloses the true LD50 for an experimental sample. Because the profile likelihood method is approximate, coverage of the confidence interval does not always correspond to its nominal value. For example, coverage falls below 95% for populations with shallow slopes and is better than 95% for populations with steep slopes. In addition, the width of the confidence interval is

limited by the dose progression chosen. Generally, no type of confidence interval would be more narrow than the dose progression.

3. Response Patterns.

Data gathered under the UDP fall into one of five animal response patterns. The five types of animal response patterns, referred to as Case 1 - Case 5 below, can be distinguished for the purpose of describing the performance of the UDP. These cases can be distinguished by looking at the experimental outcome (survival or death) as reflected in the AOT425StatPgm Data Grid or Report windows. In considering these cases, note that doses can be repeated more than once in the course of sequential dosing.

Table E.1. Outcomes of the Up-and-Down Procedure: Cases and Confidence Intervals.

Case #	Definition of Case	Approach Proposed	Possible Findings
1	**No positive dose-response association.** 1a) all animals tested in the study responded, or 1b) none responded, or 1c) the geometric mean dose is lower for animals that responded than for animals that did not respond.	LD50 cannot be calculated. Confidence interval not applicable.	Possible inferences: 1a) LD50 < lowest dose; 1b) LD50 > highest dose; 1c) reverse dose-response curve; unlikely test outcome. In case 1b, the highest dose tested is equivalent to a limit dose.
2	**Multiple partial responses.** One or more animals responded at a dose below some other dose where one or more did not respond. The conditions defining Case 1 do not hold. (The definition of Case 2 holds if there are 2 doses with partial responses, but holds in some other cases as well.)	Maximum likelihood estimate and profile likelihood computations of confidence interval are straightforward.	The LD50 can be estimated and its confidence interval calculated.
3	**No intermediate response fractions.** One or more test doses is associated with 0% response and one or more is associated with 100% response (all of the latter being greater than all of the former), and no test doses are associated with a partial response.	Lower bound = highest test dose with 0% response. Upper bound = lowest test dose with 100% response.	High confidence that the true LD50 falls between the two bounding doses. Any value of LD50 between highest dose with 0% response and lowest dose with 100% response is equally plausible.

4	**One partial response fraction, first subcase.** An intermediate partial response is observed at a single test dose. That dose is greater than doses associated with 0% response and lower than doses associated with 100% response.	The LD50 is set at the single dose showing partial response and its confidence interval is calculated using profile likelihood method.	The LD50 can be estimated and its confidence interval calculated.
5	**One partial response fraction, second subcase.** There is a single dose associated with partial response, which is either the highest test dose (with no responses at all other test doses) or the lowest test dose (with 100% response at all other test doses).	The LD50 is set at the dose with the partial response. A profile likelihood confidence interval is calculated and may be finite or infinite.	The true LD50 could be at the boundary of the testing range with more or less confidence.

APPENDIX C

August 2001 Peer Panel Meeting Review Materials

C–1 Acute Oral Toxicity: Revised UDP Test Guideline, July 12, 2001 ..... C-3

C–2 A Proposed Procedure for Calculating Confidence Intervals C-33

C–3 Description of the Acute Oral Toxicity Software Program C-109

C–4 Evaluation Guidance to the Panel for the Revised UDP..... C-111

DRAFT **rev3udp**
7/12/01

GUIDELINE FOR THE TESTING OF CHEMICALS

Acute Oral Toxicity: Revised Up-and-Down Procedure

INTRODUCTION

1. OECD guidelines for the Testing of Chemicals are periodically reviewed in the light of scientific progress or changing assessment practices. The concept of the up-and-down testing approach was first described by Dixon and Mood (1)(2)(3)(4). In 1985, Bruce proposed to use an Up-and-Down Procedure (UDP) for the determination of acute toxicity of chemicals (5). There exist several variations of the up-and-down experimental design for estimating an LD50. This guideline is based on the procedure of Bruce as adopted by ASTM in 1987 (6) and revised in 1990. A study comparing the results obtained with the UDP, the conventional LD50 test and the Fixed Dose Procedure (FDP, Guideline 420) was published in 1995 (7). Since the early papers of Dixon and Mood, papers have continued to appear in the biometrical and applied literature, examining the best conditions for use of the approach (8)(9)(10)(11). Based on the recommendations of several expert meetings in 1999, an additional revision was considered timely because: i) international agreement had been reached on harmonised LD50 cut-off values for the classification of chemical substances, ii) testing in one sex (usually females) is generally considered sufficient, and iii) there is a need to estimate confidence intervals (CI).

2. The test procedure described in this guideline is of value in minimizing the number of animals required to estimate the acute oral toxicity of a chemical. In addition to the estimation of LD50 and CIs, the test allows the observation of signs of toxicity. Revision of test guideline 425 was undertaken concurrently with two other alternatives to conventional acute oral toxicity test. Guidance on the selection of the most appropriate test method can be found in the Guidance Document on Oral Toxicity Testing (12). This Guidance Document also contains additional information on the conduct and interpretation of Guideline 420 and 423.

3. Definitions used in the context of this Guideline are set out in Annex 1.

INITIAL CONSIDERATIONS

4. All available information on the test substance should be considered by the testing laboratory prior to conducting the study. Such information will include the identity and chemical structure of the substance; its physical chemical properties; the results of any other *in vitro* or

in vivo toxicity tests on the substance or mixtures; toxicological data on structurally related substances or similar mixtures; and the anticipated use(s) of the substance. This information is useful to determine the relevance of the test for the protection of human health and the environment, and will help in the selection of an appropriate starting dose.

5. If no information is available to make a preliminary estimate of the LD50 and the slope of the dose-response curve, results of computer simulations have suggested that starting near 175 mg/kg and using half-log units (corresponding to a dose progression of 3.2) between doses will produce the best results. This starting dose should be modified if the substance is likely to be highly toxic. The half-log spacing provides for a more efficient use of animals, and increases accuracy in the prediction of the LD50 value. Because the method has a bias toward the starting dose, it is essential that initial dosing occur below the estimated LD50. (See paragraph 27 and Annex 2 for discussion of dose sequences and starting values.) However, for chemicals with large variability (i.e., shallow dose-response slopes), bias can still be introduced in the lethality estimates and the LD50 will have a large statistical error, similar to other acute toxicity methods. To correct for this, the main test includes a stopping rule keyed to properties of the estimate rather than a fixed number of test observations.

6. The method is easiest to apply to materials that produce death within one or two days. The method would not be practical to use when considerably delayed death (five days or more) can be expected.

7. Computers are used to facilitate animal-by-animal calculations that establish testing sequences and provide final estimates.

8. Moribund animals or animals obviously in pain or showing signs of severe and enduring distress shall be humanely killed, and are considered in the interpretation of the test results in the same way as animals that died on test. Criteria for making the decision to kill moribund or severely suffering animals, and guidance on the recognition of predictable or impending death are the subject of a separate OECD Guidance Document (13).

9. A limit test can be used efficiently to identify chemicals that are likely to have low toxicity.

PRINCIPLE OF THE LIMIT TEST

10. The Limit Test is a sequential test that uses a maximum of 5 animals. A test dose of up to 2000, or exceptionally 5000 mg/kg, may be used. The procedures for testing at 2000 and 5000 mg/kg are slightly different. The selection of a sequential test plan increases the statistical power and also has been made to intentionally bias the procedure towards rejection of the limit test for compounds with LD50s near the limit dose; i.e., to err on the side of safety. As with any limit test protocol, the probability of correctly classifying a compound will decrease as the actual LD50 more nearly resembles

the limit dose.

PRINCIPLE OF THE MAIN TEST

11. The main test consists of a single ordered dose progression in which animals are dosed, one at a time, at 48-hour intervals. The first animal receives a dose a step below the level of the best estimate of the LD50. If the animal survives, the dose for the next animal is increased to a factor of 3.2 times the original dose; if it dies, the dose for the next animal is decreased by a similar dose progression. (Note: 3.2 is the default factor. Paragraph 27 provides further guidance for choice of dose spacing factor.) Each animal should be observed carefully for up to 48 hours before making a decision on whether and how much to dose the next animal. That decision is based on the 48-hour survival pattern of all the animals up to that time. (See paragraphs 26 and 30 on choice of survival interval.) A combination of stopping criteria is used to keep the number of animals low while adjusting the dosing pattern to reduce the effect of a poor starting value or low slope (see paragraph 29). Dosing is stopped when one of these criteria is satisfied (see paragraphs 28 and 36), at which time an estimate of the LD50 and a CI are calculated for the test based on the status of all the animals at termination. For most applications, testing will be completed with only 4 animals after initial reversal in animal outcome. The LD50 is calculated using the method of maximum likelihood (14)(15). (See paragraphs 36 and 38.)

12. The results of the main test procedure serve as the starting point for a computational procedure to provide a CI estimate where feasible. A description of the basis for this CI is outlined in paragraph 40.

DESCRIPTION OF THE METHOD

Selection of animals species

13. The preferred rodent species is the rat although other rodent species may be used. Normally female rats are used (12). This is because literature surveys of conventional LD50 tests show that usually there is little difference in sensitivity between sexes, but in those cases where differences are observed, females are generally more sensitive (7). However, if knowledge of the toxicological or toxicokinetic properties of structurally related chemicals indicates that males are likely to be more sensitive then this sex should be used. When the test is conducted in males, adequate justification should be provided.

14. Healthy young adult animals of commonly used laboratory strains should be employed. Females should be nulliparous and non-pregnant. At the commencement of its dosing, each animal should be between 8 and 12 weeks old and its weight should fall in an interval ± 20 % of the mean initial weight of any previously dosed animals.

Housing and feeding conditions

15. The temperature in the experimental animal room should be 22EC (± 3EC). The relative humidity should be at least 30 % and preferably not exceed 70 % other than during room cleaning. Lighting should be artificial, the sequence being 12 hours light and 12 hours dark. The animals are housed individually. For feeding, conventional rodent laboratory diets may be used with an unlimited supply of drinking water.

Preparation of animals

16. The animals are randomly selected, marked to permit individual identification, and kept in their cages for at least 5 days prior to dosing to allow for acclimatisation to the laboratory conditions. As with other sequential test designs, care must be taken to ensure that animals are available in the appropriate size and age range for the entire study.

Preparation of doses

17. When necessary, the test substance is dissolved or suspended in a suitable vehicle. The use of an aqueous solution/suspension/emulsion is recommended wherever possible, followed in order of preference by a solution/suspension/emulsion in oil (e.g. corn oil) and then possibly solution in other vehicles. For vehicles other than water the toxicological characteristics of the vehicle should be known. Dosing preparations must be prepared shortly prior to administration unless the stability of the preparation over the period during which it will be used is known. Where preparation shortly before administration is not practicable and the stability of the preparation is not known, this will need to be demonstrated analytically.

17a. In general test substances should be administered in a constant volume over the range of doses to be tested by varying the concentration of the dosing preparation. Where a liquid end product or mixture is to be tested, however, the use of the undiluted test substance, i.e., at a constant concentration, may be more relevant to the subsequent risk assessment of that substance, and is a requirement of some regulatory jurisdictions. In either case, the maximum dose volume for administration must not be exceeded. The maximum volume of liquid that can be administered at one time depends on the size of the test animal. In rodents, the volume should not normally exceed 1 ml/100g of body weight; however in the case of aqueous solutions,
2 ml/100g body weight can be considered.

Administration of doses

17b The test substance is administered in a single dose by gavage using a stomach tube or a suitable intubation cannula. In the unusual circumstance that a single dose is not possible, the dose may be given

in smaller fractions over a period not exceeding 24 hours.

17c. Animals should be fasted prior to dosing (e.g., with the rat, food but not water should be withheld overnight; with the mouse, food but not water should be withheld for 3-4 hours). Following the period of fasting, the animals should be weighed and the test substance administered. The fasted body weight of each animal is determined and the dose is calculated according to the body weight. After the substance has been administered, food may be withheld for a further 3-4 hours in rats or 1-2 hours in mice. Where a dose is administered in fractions over a period of time, it may be necessary to provide the animals with food and water depending on the length of the period.

PROCEDURE

Limit test and main test

18. The limit test is primarily used in situations where the experimenter has information indicating that the test material is likely to be nontoxic, i.e., having toxicity below regulatory limit doses. Information about the toxicity of the test material can be gained from knowledge about similar tested compounds or similar tested mixtures or products, taking into consideration the identity and percentage of components known to be of toxicological significance. In those situations where there is little or no information about its toxicity, or in which the test material is expected to be toxic, the main test should be performed.

Limit test

Limit test at 2000 mg/kg

19. Dose one animal at the test dose. If the animal dies, conduct the main test to determine the LD50. If the animal survives, dose four additional animals, one at a time. If an animal unexpectedly dies late in the study, and there are other survivors, it is appropriate to stop dosing and observe all animals to see if other animals will also die during a similar observation period (see paragraph 26 for initial observation period). Late deaths should be counted the same as other deaths. The results are evaluated as follows (O=survival, X=death).

20. The LD50 is less than the test dose (2000 mg/kg) when three or more animals die.

O XO XX
O OX XX
O XX OX
O XX XU (U can be O or X)

If a third animal dies, conduct the main test.

21. The LD50 is greater than the test dose (2000 mg/kg) when three or more animals survive.

O OO OO
O OO XO
O OO OX
O OO XX
O XO XO
O XO OU (U can be O or X)
O OX XO
O OX OU (U can be O or X)
O XX OO

Limit Test at 5000 mg/kg

22. Exceptionally, and only when justified by specific regulatory needs, the use of a dose at 5000 mg/kg may be considered. Recognizing the need to protect animal welfare, testing of animals in class 5 ranges is discouraged and should only be considered when there is a strong likelihood that results of such a test have a direct relevance for protecting human or animal health or the environment.

23. Dose one animal at the test dose. If the animal dies, conduct the main test to determine the LD50. If the animal survives, dose two additional animals. If both animals survive, the LD50 is greater than the limit dose and the test is terminated (i.e. carried to full 14-day observation without dosing of further animals). If one or both animals die, then dose an additional two animals, one at a time. If an animal unexpectedly dies late in the study, and there are other survivors, it is appropriate to stop dosing and observe all animals to see if other animals will also die during a similar observation period (see paragraph 10 for initial observation period). Late deaths should be counted the same as other deaths. The results are evaluated as follows (O=survival, X=death, and U=Unnecessary).

24. The LD50 is less than the test dose (5000 mg/kg) when three or more animals die.

O XO XX
O OX XX
O XX OX
O XX XU (U can be O or X, the dosing of the 5th animal is not necessary)

25. The LD50 is greater than the test dose (5000 mg/kg) when three or more animals survive.

O OO UU (U can be O or X, the dosing of the 4th and 5th animal is not necessary)
O XO XO

O XO OU (U can be O or X, the dosing of the 5th animal is not necessary)
O OX XO
O OX OU (U can be O or X, the dosing of the 5th animal is not necessary)
O XX OO

Main test

26. Single animals are dosed in sequence usually at 48 h intervals. However, the time intervals between dosing is determined by the onset, duration, and severity of toxic signs. Treatment of an animal at the next dose should be delayed until one is confident of survival of the previously dosed animal. The time interval may be adjusted as appropriate, e.g., in case of inconclusive response. The test is simpler to implement when a single time interval is used for making sequential dosing decisions. Nevertheless, it is not necessary to recalculate dosing or likelihood-ratios if the time interval changes midtest. For selecting the starting dose, all available information, including information on structurally related substances and results of any other toxicity tests on the test material, should be used to approximate the LD50 as well as the slope of the dose-response curve.

27. The first animal is dosed a step below the toxicologist's best estimate of the LD50. If the animal survives, the second animal receives a higher dose. If the first animal dies or appears moribund, the second animal receives a lower dose. The dose progression factor should be chosen to be the antilog of 1/(the estimated slope of the dose-response curve) and should remain constant throughout testing. When there is no information on the slope of the substance to be tested, a dose progression factor of 3.2 is used. Using the default progression factor, doses would be selected from the sequence 1.75, 5.5, 17.5, 55, 175, 550, 2000 (or 1.75, 5.5, 17.5, 55, 175, 550, 1750, 5000 for specific regulatory needs). If no estimate of the substance's lethality is available, dosing should be initiated at 175 mg/kg. In most cases, this dose is sublethal and therefore serves to reduce the level of pain and suffering. If animal tolerances to the chemical are expected to be highly variable (i.e., slopes are expected to be less than 2.5), consideration should be given to increasing the dose progression factor beyond the default 0.5 on a log dose scale (i.e., 3.2 progression factor) prior to starting the test. (Annex 2 includes a table of dose progressions for whole number slopes ranging from 1 to 8 with starting dose 175 mg/kg.)

28. Dosing continues depending on the fixed-time interval (e.g., 48-hour) outcomes of all the animals up to that time. The testing stops when one of the following stopping criteria first is met:

(a) 3 consecutive animals survive at the upper bound;
(b) 5 reversals occur in any 6 consecutive animals tested;
(c) at least 4 animals have followed the first reversal and the specified likelihood-ratios exceed the critical value. (See paragraph 39 and Annex 3. Calculations are made at each dosing, following the fourth animal after the first reversal.).

For a wide variety of combinations of LD50 and slopes, stopping rule (c) will be satisfied with 4 to 6 animals after the test reversal. In some cases for chemicals with shallow slope dose-response curves, additional animals (up to a total of fifteen tested may be needed).

29. When the stopping criteria have been attained, the estimated LD50 should be calculated from the animal outcomes at test termination using the method described in paragraphs 35 and 36.

30. Moribund animals killed for humane reasons are considered in the same way as animals that died on test. If an animal unexpectedly dies late in the study and there are other survivors at that dose or above, it is appropriate to stop dosing and observe all animals to see if other animals will also die during a similar observation period. If subsequent survivors also die, AND it appears that all dose levels exceed the LD50 it would be most appropriate to start the study again beginning at least two steps below the lowest dose with deaths (and increasing the observation period) since the technique is most accurate when the starting dose is below the LD50. If subsequent animals survive at or above the dose of the animal that dies, it is not necessary to change the dose progression since the information from the animal that has now died will be included into the calculations as a death at a lower dose than subsequent survivors, pulling the LD50 down.

Observations

31. Animals are observed individually at least once during the first 30 minutes after dosing, periodically during the first 24 hours (with special attention given during the first 4 hours), and daily thereafter, for a total of 14 days, except where they need to be removed from the study and humanely killed for animal welfare reasons or are found dead. However, the duration of observation should not be fixed rigidly. It should be determined by the toxic reactions and time of onset and length of recovery period, and may thus be extended when considered necessary. The times at which signs of toxicity appear and disappear are important, especially if there is a tendency for toxic signs to be delayed (16). All observations are systematically recorded with individual records being maintained for each animal.

32. Additional observations will be necessary if the animals continue to display signs of toxicity. Observations should include changes in skin and fur, eyes and mucous membranes, and also respiratory, circulatory, autonomic and central nervous systems, and somatomotor activity and behaviour pattern. Attention should be directed to observations of tremors, convulsions, salivation, diarrhoea, lethargy, sleep and coma. The principles and criteria summarised in the Humane Endpoints Guidance Document (13) should be taken into consideration. Animals found in a moribund condition and animals showing severe pain and enduring signs of severe distress should be humanely killed. When animals are killed for humane reasons or found dead, the time of death should be recorded as precisely as possible.

Body weight

33. Individual weights of animals should be determined shortly before the test substance is administered and at least weekly thereafter. Weight changes should be calculated and recorded. At the end of the test surviving animals are weighed and then humanely killed.

Pathology

34. All animals (including those which die during the test or are removed from the study for animal welfare reasons) should be subjected to gross necropsy. All gross pathological changes should be recorded for each animal. Microscopic examination of organs showing evidence of gross pathology in animals surviving 24 or more hours after the initial dosing may also be considered because it may yield useful information.

DATA AND REPORTING

Data

35. Individual animal data should be provided. Additionally, all data should be summarized in tabular form, showing for each test dose the number of animals used, the number of animals displaying signs of toxicity (16), the number of animals found dead during the test or killed for humane reasons, time of death of individual animals, a description and the time course of toxic effects and reversibility, and necropsy findings. A rationale for the starting dose and the dose progression and any data used to support this choice should be provided.

Calculation of LD50 for the main test

36 The LD50 is calculated using the maximum likelihood method (14)(15)**,** except in the exceptional cases described in paragraph 37. The following statistical details may be helpful in implementing the maximum likelihood calculations suggested (with an assumed *sigma*). All deaths, whether immediate or delayed or humane kills, are incorporated for the purpose of the maximum likelihood analysis. Following Dixon (4), the likelihood function is written as follows:

$$L = L_1 L_2 \ldots L_n ,$$

where

L is the likelihood of the experimental outcome, given *mu* and *sigma*, and n the total number of animals tested.

$L_i = 1 - F(Z_i)$ if the ith animal survived, or
$L_i = F(Z_i)$ if the ith animal died,

where

F = cumulative standard normal distribution,
$Z_i = [\log(d_i) - mu\ ] / sigma$
d_i = dose given to the ith animal, and
sigma = standard deviation in log units of dose (which is not the log standard deviation).

When identifying the maximum of the likelihood L to get an estimate of the true LD50, *mu* is set to equal log LD50, and automated calculations solve for it (see paragraph 39).

An estimate of *sigma* of 0.5 is used unless a better generic or case-specific value is available.

37. Under some circumstances, statistical computation will not be possible or will likely give erroneous results. Special means to determine/report an estimated LD50 are available for these circumstances as follows:

(a) If testing stopped based on criterion (a) in paragraph 28 (i.e., a boundary dose was tested repeatedly), or if the upper bound dose ended testing, then the LD50 is reported to be above the upper bound. Classification is completed on this basis.
(b) If all the dead animals have higher doses than all the live animals (or if all live animals have higher doses than all the dead animals, although this is practically unlikely), then the LD50 is between the doses for the live and the dead animals. These observations give no further information on the exact value of the LD50. Still, a maximum likelihood LD50 estimate can be made provided there is a value for *sigma*. Stopping criterion (b) in paragraph 28 describes one such circumstance.
(c) If the live and dead animals have only one dose in common and all the other dead animals have higher doses and all the other live animals lower doses, or vice versa, then the LD50 equals their common dose. If a closely related substance is tested, testing should proceed with a smaller dose progression.

If none of the above situations occurs, then the LD50 is calculated using the maximum likelihood method.

38. Maximum likelihood calculation can be performed using either SAS (14)(e.g., PROC NLIN) or BMDP (15)(e.g., program AR) computer program packages as described in Appendix 1D in Reference 3. Other computer programs may also be used. Typical instructions for these packages are given in appendices to the ASTM Standard E 1163-87 (6). [The *sigma* used in the BASIC program in

(6) will need to be edited to reflect the parameters of this OECD 425 Guideline.] The program's output is an estimate of log(LD50) and its standard error.

39. The likelihood-ratio stopping rule (c) in paragraph 28 is based on three measures of test progress, that are of the form of the likelihood in paragraph 36, with different values for *mu*. Comparisons are made after each animal tested after the sixth that does not already satisfy criterion (a) or (b) of paragraph 28. The equations for the likelihood-ratio criteria are provided in Annex 3. These comparisons are most readily performed in an automated manner and can be executed repeatedly, for instance, by a spreadsheet routine such as that also provided in Annex 3. If the criterion is met, testing stops and the LD50 can be calculated by the maximum likelihood method.

Computation of confidence interval

40. Following the main test and estimated LD50 calculation, it may be possible to compute interval estimates for the LD50 at specified confidence using a profile-likelihood-based computational procedure. Such an interval utilizes information from the doses where accumulated response was neither 0% nor 100% (intermediate doses). Instead of employing an assumed *sigma*, however, the procedure identifies bounds on LD50 estimates from a ratio of likelihood functions optimized over *sigma* (profile likelihoods). Procedures are also included for certain circumstances where no intermediate doses exist (for instance, when testing has proceeded through a wide range of doses with no reversal or where doses are so widely spaced that each animal provides a reversal). Implementing this set of procedures requires specialized computation which is either by use of a dedicated program to be available from OECD or developed following technical details available from OECD.

Report

41. The test report must include the following information:

Test substance:

- physical nature, purity and physicochemical properties (including isomerisation);
- identification data.

Vehicle (if appropriate):

- justification for choice of vehicle, if other than water.

Test animals:

- species/strain used;
- microbiological status of the animals, when known;

- number, age and sex of animals;
- rationale for use of males instead of females;
- source, housing conditions, diet, etc.;
- individual weights of animals at the start of the test, at day 7, and at day 14.

Test conditions:

- rationale for initial dose level selection, dose progression factor and for follow-up dose levels;
- details of test substance formulation;
- details of the administration of the test substance;
- details of food and water quality (including diet type/source, water source).

Results:

- body weight/body weight changes;
- tabulation of response data by sex (if both sexes are used) and dose level for each animal (i.e., animals showing signs of toxicity including nature, severity, duration of effects, and mortality);
- time course of onset of signs of toxicity and whether these were reversible for each animal;
- necropsy findings and any histopathological findings for each animal, if available;
- slope of the dose-response curve (when determined);
- LD50 data;
- statistical treatment of results (description of computer routine used and spreadsheet tabulation of calculations)

Discussion and interpretation of results.

Conclusions.

LITERATURE

(1) Dixon, W.J. and A.M. Mood. (1948). A Method for Obtaining and Analyzing Sensitivity Data. J. Amer. Statist. Assoc., 43, 109-126.

(2) Dixon, W.J. The Up-and-Down Method for Small Samples (1965). J. Amer. Statist. Assoc. 60, 967-978.

(3) Dixon, W.J. (1991). Staircase Bioassay: The Up-and-Down Method. Neurosci. Biobehav. Rev., 15, 47-50.

(4) Dixon, W.J. (1991) Design and Analysis of Quantal Dose-Response Experiments (with Emphasis on Staircase Designs). Dixon Statistical Associates, Los Angeles CA, USA.

(5) Bruce, R.D. (1985). An Up-and-Down Procedure for Acute Toxicity Testing. Fundam. Appl. Tox., 5, 151-157.

(6) ASTM (1987). E 1163-87, Standard Test Method for Estimating Acute Oral Toxicity in Rats. American Society for Testing and Materials, Philadelphia Pa, USA.

(7) Lipnick, R.L., J.A. Cotruvo, R.N. Hill, R.D. Bruce, K.A. Stitzel, A.P. Walker, I. Chu, M. Goddard, L. Segal, J.A. Springer, and R.C. Myers. (1995). Comparison of the Up-and-Down, Conventional LD_{50} and Fixed Dose Acute Toxicity Procedures. Fd. Chem. Toxicol., 33, 223-231.

(8) Choi, S.C. (1990). Interval estimation of the LD_{50} based on an up-and-down experiment. Biometrics 46, 485-492.

(9) Vågerö, M. and R. Sundberg. (1999). The distribution of the maximum likelihood estimator in up-and-down experiments for quantal dose-response data. J. Biopharmaceut. Statist. 9(3), 499-519.

(10) Hsi, B.P. (1969). The multiple sample up-and-down method in bioassay. J. Amer. Statist. Assoc. 64, 147-162.

(11) Noordwijk, A.J. van and J. van Noordwijk. (1988). An accurate method for estimating an approximate lethal dose with few animals, tested with a Monte Carlo procedure. Arch. Toxicol. 61, 333-343.

(12) OECD (2000) Guidance Document on Acute Oral Toxicity. Environmental Health and Safety Monograph Series on Testing and Assessment No. 24.

(13) OECD. (2000). Revised Draft Guidance Document on the Recognition, Assessment and Use of Clinical Signs as Humane Endpoints for Experimental Animals Used in Safety Evaluation.

(14) SAS Institute Inc. (1990). SAS/STAT® User's Guide. Version 6, Fourth Ed. or later. Cary, NC, USA.

(15) BMDP Statistics Software, Inc. (1990). BMDP Statistical Software Manual. W.J. Dixon, Chief Ed. 1990 rev. or later. University of California Press, Berkeley, CA, USA.

(16) Chan, P.K. and A.W. Hayes. (1994). Chap. 16. Acute Toxicity and Eye Irritancy.

Principles and Methods of Toxicology. Third Edition. A.W. Hayes, Editor. Raven Press, Ltd., New York, USA.

ANNEX 1

DEFINITIONS

Acute oral toxicity is the adverse effects occurring within a short time of oral administration of a single dose of a substance or multiple doses given within 24 hours.

Delayed death means that an animal does not die or appear moribund within 48 hours but dies later during the 14-day observation period.

Dose is the amount of test substance administered. Dose is expressed as weight (g, mg) or as weight of test substance per unit weight of test animal (e.g. mg/kg).

Dose progression factor, sometimes termed a dose spacing factor, refers to the multiple by which a dose is increased (i.e., the dose progression) when an animal survives or the divisor by which it is decreased when an animal dies.

LD50 (median lethal dose), oral, is a statistically derived single dose of a substance that can be expected to cause death in 50 per cent of animals when administered by the oral route. The LD50 value is expressed in terms of weight of test substance per unit weight of test animal (mg/kg).

Limit dose refers to a dose at an upper limitation on testing (2000-5000 mg/kg).

Moribund status of an animal refers to being in a state of dying or inability to survive, even if treated.

Nominal sample size refers to the total number of tested animals**,** reduced by one less than the number of like responses at the beginning of the series, or by the number of tested animals up to but not including the pair that creates the first reversal. For example, for a series where X and O indicate opposite animal outcomes (for instance, X could be dies within 48 hours and O survives) in a pattern as follows: OOOXXOXO, we have the total number of tested animals (or sample size in the conventional sense) as 8 and the nominal sample size as 6. This particular example shows 4 animals following a reversal. It is important to note whether a count in a particular part of the guideline refers to the nominal sample size or to the total number tested. For example, the maximum actual number tested is 15. When testing is stopped based on that basis, the nominal sample size will be less than or equal to 15. Members of the nominal sample start with the (r-1)st animal (the animal before the second in the reversal pair) (see reversal below).

Probit is an abbreviation for the term "probability integral transformation" and a probit dose-response model permits a standard normal distribution of expected responses (i.e., one centered to its mean and scaled to its standard deviation, *sigma*) to doses (typically in a logarithmic scale) to be analyzed as if it were a straight line with slope the reciprocal of sigma. A standard normal lethality distribution is

symmetric; hence, its mean is also its true LD50 or median response.

Reversal is a situation where nonresponse is observed at some dose, and a response is observed at the next dose tested, or vice versa (i.e., response followed by nonresponse). Thus, a reversal is created by a pair of responses. The first such pair occurs at animals numbered r-1 and r.

Sigma is the standard deviation of a log normal curve describing the range of tolerances of test subjects to the chemical (where a subject is expected capable of responding if the chemical dose exceeds the subject's tolerance). The estimated *sigma* provides an estimate of the variation among test animals in response to a full range of doses. See slope and probit.

Slope (of the dose-response curve) is a value related to the angle at which the dose response curve rises from the dose axis. In the case of probit analysis, when responses are analyzed on a probit scale against dose on a log scale this curve will be a straight line and the slope is the reciprocal of *sigma*, the standard deviation of the underlying test subject tolerances, which are assumed to be normally distributed. See probit and *sigma*.

Stopping rule is used in this guideline synonymously with 1) a specific stopping criterion and 2) the collection of all criteria determining when a testing sequence terminates. In particular, for the main test, stopping rule is used in paragraph 5 as a shorthand for the criterion that relies on comparison of ratios to a critical value.

ANNEX 2

DOSING PROCEDURE

Dose Sequence for Main Test

1. Up-and-Down Dosing Procedure. For each run, animals are dosed, one at a time, usually at 48-hour intervals. The first animal receives a dose a step below the level of the best estimate of the LD50. This selection reflects an adjustment for a tendency to bias **away from the LD50 in the direction of** the initial starting dose in the final estimate (see paragraph 5). The overall pattern of outcomes is expected to stabilize as dosing is adjusted for each subsequent animal. Paragraph 3 below provides further guidance for choice of dose spacing factor).

2. Default Dose Progression. Once the starting dose and dose spacing are decided, the toxicologist should list all possible doses including the upper bound (usually 2000 or 5000 mg/kg). Doses that are close to the upper bound should be removed from the progression. The stepped nature of the TG 425 design provides for the first few doses to function as a self-adjusting sequence. Because of the tendency for positive bias, in the event that nothing is known about the substance, a starting dose of 175 mg/kg is recommended. If the default procedure is to be used for the main test, dosing will be initiated at 175 mg/kg and doses will be spaced by a factor of 0.5 on a log dose scale. The doses to be used include 1.75, 5.5, 17.5, 55, 175, 550, 2000 or, for specific regulatory needs, 1.75, 5.5, 17.5, 55, 175, 550, 1750, 5000. For certain highly toxic substances, the dosing sequence may need to be extended to lower values.

3. In the event a dose progression factor other than the default is deemed suitable, Table 1 provides dose progressions for whole number multiples of slope, from 1 to 8.

Table 1 Dose Progressions for OECD Guideline 425
Choose a Slope and Read Down the Column
All doses in mg/kg bw

Slope =	1	2	3	4	5	6	7	8
	0.175*	0.175*	0.175*	0.175*	0.175*	0.175*	0.175*	0.175*
							0.243	0.233
					0.28	0.26		
				0.31			0.34	0.31
			0.38			0.38		
								0.41
					0.44		0.47	
		0.55		0.55		0.55		0.55
					0.70		0.65	
								0.74
			0.81			0.81		
				0.98			0.91	0.98
					1.10	1.19		
							1.26	1.31
	1.75	1.75	1.75	1.75	1.75	1.75	1.75	1.75
							2.43	2.33
					2.8	2.6		
				3.1			3.4	3.1
			3.8			3.8		
					4.4			4.1
							4.7	
		5.5		5.5		5.5		5.5
					7.0		6.5	
								7.4
			8.1			8.1		
				9.8			9.1	9.8
					11.0	11.9		
							12.6	13.1
	17.5	17.5	17.5	17.5	17.5	17.5	17.5	17.5
							24.3	23.3
					28	26		
				31			34	31

Table 1 continued

		38			38		
				44			41
						47	
	55		55		55		55
						65	
				70			74
		81			81		
			98			91	98
				110	119		
						126	131
175	175	175	175	175	175	175	175
						243	233
				280	260		
			310			340	310
		380			380		
				440			410
						470	
	550		550		550		550
						650	
				700			740
		810			810		
			980			910	980
				1100	1190		
						1260	1310
1750	1750	1750	1750	1750	1750	1750	1750
						2430	2330
				2800	2600		
			3100				3100
					3800	3400	
							4100
5000	5000	5000	5000	5000	5000	5000	5000

* If lower doses are needed, continue progressions to a lower dose

ANNEX 3

COMPUTATIONS FOR THE LIKELIHOOD-RATIO STOPPING RULE

As described in Guideline paragraph 28, the main test may be completed on the basis of the first of four stopping criteria to occur. Tables 2-5 illustrate examples where testing has started with no information, so the recommended default starting value, 175 mg/kg, and the recommended default dose progression factor, 3.2, have been used.

Table 2 shows how the main test would stop if 3 animals have survived at the limit dose of 2000 mg/kg; Table 3 shows a similar situation when the limit dose of 5000 mg/kg is used. (These illustrate situations where a Limit Test was not thought appropriate *a priori*.) Table 4 shows how a particular sequence of 5 reversals in 6 tested animals could occur and allow test completion. Finally, Table 5 illustrates a situation several animals into a test, where neither criterion (a) nor criterion (b) has been met, a reversal of response has occurred followed by 4 tested animals, and, consequently, criterion (c) must be evaluated as well.

Criterion (c) calls for a likelihood-ratio stopping rule to be evaluated after testing each animal, starting with the fourth tested following the reversal. Three "measures of test progress" are calculated. Technically, these measures of progress are likelihoods, as recommended for the maximum-likelihood estimation of the LD50. The procedure is closely related to calculation of a CI by a likelihood-based procedure.

The basis of the procedure is that when enough data have been collected, a point estimate of the LD50 should be more strongly supported than values above and below the point estimate, where statistical support is quantified using likelihood. Therefore three likelihood values are calculated: a likelihood at an LD50 point estimate (called the rough estimate or dose-averaging estimate in the example), a likelihood at a value below the point estimate, and a likelihood at a value above the point estimate. Specifically, the low value is taken to be the point estimate divided by 2.5 and the high value is taken to be the point estimate multiplied by 2.5.

The likelihood values are compared by calculating ratios of likelihoods, and then determining whether these likelihood-ratios (LR) exceed a critical value. Testing stops when the ratio of the likelihood for the point estimate exceeds each of the other likelihoods by a factor of 2.5, which is taken to indicate relatively strong statistical support for the point estimate. Therefore two likelihood-ratios (LRs) are calculated, a ratio of likelihoods for the point estimate and the point estimate divided by 2.5, and a ratio for the point estimate and the estimate times 2.5.

The calculations are easily performed in any spreadsheet with normal probability functions. The calculations are illustrated in Table 5, which is structured to promote spreadsheet implementation. The

computation steps are illustrated using an example where the upper limit dose is 5000 mg/kg, but the computational steps are carried out in the same fashion when the upper boundary dose is 2000 mg/kg. Empty spreadsheets preprogrammed with the necessary formulas are available for direct downloading on the OECD and EPA web sites.

Hypothetical example using an upper limit dose of 5000 mg/kg (Table 5)

In the hypothetical example utilizing an upper boundary dose of 5000 mg/kg, the LR stopping criterion was met after nine animals had been tested. The first "reversal" occurred with the 3rd animal tested. The LR stopping criterion is checked when four animals have been tested following the reversal. In this example, the fourth animal tested following the reversal is the seventh animal actually tested. Therefore, for this example, the spreadsheet calculations are only needed after the seventh animal had been tested and the data could be entered at that time. Subsequently, the LR stopping criterion would have been checked after testing the seventh animal, the eighth animal, and the ninth. The LR stopping criterion is first satisfied after the ninth animal is tested in this example.

A. Enter the dose-response information animal by animal.

Column 1. Steps are numbered 1-15. No more than 15 animals may be tested.
Column 2. Place an I in this column as each animal is tested.
Column 3. Enter the dose received by the ith animal.
Column 4. Indicate whether the animal responded (shown by an X) or did not respond (shown by an O).

B. The nominal and actual sample sizes.

The nominal sample consists of the two animals that represent the first reversal (here the second and third animals), plus all animals tested subsequently. Here, Column 5 indicates whether or not a given animal is included in the nominal sample.

- The nominal sample size (nominal *n*) appears in Row 16. This is the number of animals in the nominal sample. In the example, nominal *n* is 8.
- The actual number tested appears in Row 17.

C. Rough estimate of the LD50.

The geometric mean of doses for the animals in the current nominal sample is used as a rough estimate of the LD50 from which to gauge progress. In the table, this is called the "dose-averaging estimator." It is updated with each animal tested. This average is restricted to the nominal sample in order to allow for a poor choice of initial test dose, which could generate either an initial string of responses or an initial string of nonresponses. (However, the results for all animals are used in the likelihood calculations for final LD50 calculation below.) Recall that the geometric mean of *n* numbers is the product of the *n*

numbers, raised to a power of $1/n$.

- The dose-averaging estimate appears in Row 18 (e.g., $(175 * 550 * ... * 1750)^{1/8} = 1292.78$).

- Row 19 shows the logarithm (base 10) of the value in Row 18 (e.g., $\log_{10} 1292.8 = 3.112$).

D. <u>Likelihood for the rough LD50 estimate.</u>

"Likelihood" is a statistical measure of how strongly the data support an estimate of the LD50 or other parameter. Ratios of likelihood values can be used to compare how well the data support different estimates of the LD50.

In Column 8 calculate the likelihood for Step C's rough LD50 estimate. The likelihood (Row 21) is the product of likelihood contributions for individual animals (see Guideline paragraph 36). The likelihood contribution for the ith animal is denoted L_i.

Column 7. Enter the estimate of the probability of response at dose d_i, denoted P_i. P_i is calculated from a dose-response curve. Note that the parameters of a probit dose-response curve are the slope and the LD50, so values are needed for each of those parameters. For the LD50 the dose-averaging estimate from Row 18 is used. For the slope in this example the default value of 2 is used. The following steps may be used to calculate the response probability P_i.

1. Calculate the base-10 log of dose d_i (Column 6).

2. For each animal calculate the z-score, denoted Z_i (not shown in the table), using the formulae

$$sigma = 1 / slope,$$
$$Z_i = (\log_{10}(d_i) - \log_{10}(LD50)) / sigma$$

For example, for the first animal (Row 1),

$$sigma = 1 / 2$$
$$Z_1 = (2.243 - 3.112) / 0.500 = -1.738$$

3. For the ith dose the estimated response probability is

$$P_i = F(Z_i)$$

where F denotes the cumulative distribution function for the standard normal distribution (i.e., the normal distribution with mean 0 and variance 1).

For example (Row 1),

$$P_1 = F(-1.738) = 0.0412$$

The function F (or something very close) is ordinarily what is given for the normal distribution in statistical tables, but the function is also widely available as a spreadsheet function. It is available under different names, for example the @NORMAL function of Lotus 1-2-3 (1) and the @NORMDIST function in Excel (2). To confirm that you have used correctly the function available in your software, you may wish to verify familiar values such as $F(1.96)$. 0.975 or $F(1.64)$. 0.95.

Column 8. Calculate the natural log of the likelihood contribution ($\ln(L_i)$). L_i is simply the probability of the response that actually was observed for the ith animal:

responding animals: $\ln(L_i) = \ln(P_i)$
non-responding animals: $\ln(L_i) = \ln(1 - P_i)$

Note that here the natural logarithm (ln) is used, whereas elsewhere the base-10 (common) logarithm was used. These choices are what are ordinarily expected in a given context.

The steps above are performed for each animal. Finally:

Row 20: Sum the log-likelihood contributions in Column 8.
Row 21: Calculate the likelihood by applying the exp function applied to the log-likelihood value in Row 20 (e.g., $\exp(-3.389) = e^{-3.389} = 0.0337$).

E. <u>Calculate likelihoods for two dose values above and below the rough estimate.</u>

If the data permit a precise estimate, then one expects the likelihood should be high if the estimate is a reasonable estimate of the LD50, relative to likelihoods for values distant from this estimate. Compare the likelihood for the dose-averaging estimate (1292.8, Row 18) to values differing by a factor of 2.5 from that value (i.e., to 1292.8*2.5 and 1292.8/2.5). The calculations (displayed in Columns 9-12) are carried out in a fashion similar to those described above, except that the values 517.1 (=1292.8/2.5) and 3232.0 (=1292.8*2.5) have been used for the LD50, instead of 1292.8. The likelihoods and log-likelihoods are displayed in Rows 20-21.

F. <u>Calculate likelihood-ratios.</u>

The three likelihood values (Row 21) are used to calculate two likelihood-ratios (Row 22). A likelihood-ratio is used to compare the statistical support for the estimate of 1292.8 to the support for each of the other values, 517.1 and 3232.0. The two likelihood-ratios are therefore:

LR1 = [likelihood of 1292.8] / [likelihood of 517.1]
= 0.0337 / 0.0080
= 4.21

and

LR2 = [likelihood of 1292.8] / [likelihood of 3232.0]
= 0.0337 / 0.0098
= 3.44

G. Determine if the likelihood-ratios exceed the critical value.

High likelihood-ratios are taken to indicate relatively high support for the point estimate of the LD50. Both of the likelihood-ratios calculated in Step F (4.21 and 3.44) exceed the critical likelihood-ratio, which is 2.5. Therefore the LR stopping criterion is satisfied and testing stops. This is indicated by a TRUE in Row 24 and a note at the top of the example spreadsheet that the LR criterion is met.

LITERATURE

(1) Lotus Development Corporation. (1999). Lotus® 1-2-3. Version 9.5, Millennium Edition. Cambridge, MA, USA.

(2) Microsoft Corporation. (1985-1997). Microsoft® Excel. Version 5.0 or later. Seattle, WA, USA.

Table 2. Example of stopping criterion (a) using 2000 mg/kg.

Stop after animal #5 because 3 animals survive at limit of 2000 mg/kg (#3-#5).

1	2	3	4	5	6	7	8	9	10	11	12
Step	(I)nclude; (E)xclude	Dose	(X)response (O)non-resp. OK	Included in nominal *n*	log10 Dose	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)	LD50 = Prob. of response	#DIV/0! likelihood contribn. (ln *Li*)
1	I	175	O	no	2.2430	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
2	I	550	O	no	2.7404	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
3	I	2000	O	no	3.3010	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
4	I	2000	O	no	3.3010	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
5	I	2000	O	no	3.3010	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
6	E				-						
7	E				-						
8	E				-	-	-	-	-	-	-
9	E				-	-	-	-	-	-	-
10	E				-	-	-	-	-	-	-
11	E				-	-	-	-	-	-	-
12	E				-	-	-	-	-	-	-
13	E							-	-	-	-
14	E							-	-	-	-
15	E							-	-	-	-
Nominal Sample size =				0							
Actual number tested =				5							

Ignore all calculation cells. No reversal in direction of response.

Maximum Likelihood Calculations cannot be completed. LD50 is greater than 2000 mg/kg.

Calculated maximum likelihood estimate of LD50 = none

Table 3. Example of stopping criterion (a) using 5000 mg/kg.

Stop after animal #6 because 3 animals survive at limit of 5000 mg/kg (#4-#6).

1	2	3	4	5	6	7	8	9	10	11	12
Step	(I)nclude; (E)xclude	Dose	(X)response (O)non-resp.	Included in nominal *n*	log10 Dose	LD50 =	#DIV/0!	LD50 =	#DIV/0!	LD50 =	#DIV/0!
			OK			Prob. of response	likelihood contribn. (ln *Li*)	Prob. of response	likelihood contribn. (ln *Li*)	Prob. of response	likelihood contribn. (ln *Li*)
1	I	175	O	no	2.2430	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
2	I	550	O	no	2.7404	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
3	I	1750	O	no	3.2430	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
4	I	5000	O	no	3.6990	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
5	I	5000	O	no	3.6990	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
6	I	5000	O	no	3.6990	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!	#DIV/0!
7	E				-						
8	E				-						
9	E				-	-	-	-	-	-	-
10	E				-	-	-	-	-	-	-
11	E				-	-	-	-	-	-	-
12	E				-				-	-	-
13	E				-				-	-	-
14	E				-				-	-	-
15	E				-	-	-	-	-	-	-
Nominal Sample size =				0							
Actual number tested =				6							

Ignore all calculation cells. No reversal in direction of response.

Maximum Likelihood Calculations cannot be completed. LD50 is greater than 5000 mg/kg.

Calculated maximum likelihood estimate of LD50 = none

Table 4. Example of stopping criterion (b)

Stop after animal #7 because 5 reversals in 6 consecutive animals tested (#2-#7).

1	2	3	4	5	6	7	8	9	10	11	12
Step	(I)nclude; (E)xclude	Dose	(X)response (O)non-resp.	Included in nominal *n*	log10 Dose	LD50 =	31.0	LD50 =	12.4	LD50 =	77.6
						Prob. of response	likelihood contribn. (ln *Li*)	Prob. of response	likelihood contribn. (ln *Li*)	Prob. of response	likelihood contribn. (ln *Li*)
			OK								
1	I	175	X	no	2.2430	0.9335	-0.0688	0.9892	-0.0108	0.7602	-0.2742
2	I	55	X	yes	1.7404	0.6905	-0.3703	0.9020	-0.1031	0.3826	-0.9607
3	I	17.5	O	yes	1.2430	0.3095	-0.3703	0.6174	-0.9607	0.0980	-0.1031
4	I	55	X	yes	1.7404	0.6905	-0.3703	0.9020	-0.1031	0.3826	-0.9607
5	I	17.5	O	yes	1.2430	0.3095	-0.3703	0.6174	-0.9607	0.0980	-0.1031
6	I	55	X	yes	1.7404	0.6905	-0.3703	0.9020	-0.1031	0.3826	-0.9607
7	I	17.5	O	yes	1.2430	0.3095	-0.3703	0.6174	-0.9607	0.0980	-0.1031
8	E				-	-	-	-	-	-	-
9	E				-	-	-	-	-	-	-
10	E				-	-	-	-	-	-	-
11	E				-	-	-	-	-	-	-
12	E				-	-	-	-	-	-	-
13	E				-	-	-	-	-	-	-
14	E				-	-	-	-	-	-	-
15	E				-	-	-	-	-	-	-
Nominal Sample size =				6							
Actual number tested =				7							
Dose-averaging estimator				31.02							
log10 =				1.492							
log-likelihood sums:							-2.2906		-3.2021		-3.4655
likelihoods:							0.1012		0.0407		0.0313
likelihood ratios:									2.4880		3.2378
Individual ratios exceed critical value?				critical=	2.5				FALSE		TRUE
Both ratios exceed critical value?									FALSE		
Calculated maximum likelihood estimate of LD50 =					29.6						

Automated calculation; not relevant to this case.

Final estimate obtained from Maximum Likelihood Calculations

Table 5. Example of stopping criterion (c)

Stop when LR criterion is first met, here at animal #9.
Check LR criterion starting at animal #6.

Assumed slope	2	sigma =	0.5

Result:	The LR criterion is met

Parameters of convergence criterion	
critical LR	2.5
factor of LD50	2.5

1	2	3	4	5	6		7	8	9	10	11	12
Step	(I)nclude; (E)xclude	Dose	(X)response (C)non-resp. OK	Included in nominal *n*	log10 Dose	Contrib.to DAE	LD50 = Prob. of response	1292.8 likelihood contribn. (ln *Li*)	LD50 = Prob. of response	517.1 likelihood contribn. (ln *Li*)	LD50 = Prob. of response	3232.0 likelihood contribn. (ln *Li*)
1	I	175	O	no	2.2430	0.0000	0.0412	-0.0421	0.1733	-0.1903	0.0057	-0.0057
2	I	550	O	yes	2.7404	2.7404	0.2289	-0.2600	0.5214	-0.7368	0.0620	-0.0640
3	I	1750	X	yes	3.2430	3.2430	0.6037	-0.5046	0.8552	-0.1564	0.2971	-1.2138
4	I	550	O	yes	2.7404	2.7404	0.2289	-0.2600	0.5214	-0.7368	0.0620	-0.0640
5	I	1750	X	yes	3.2430	3.2430	0.6037	-0.5046	0.8552	-0.1564	0.2971	-1.2138
6	I	550	O	yes	2.7404	2.7404	0.2289	-0.2600	0.5214	-0.7368	0.0620	-0.0640
7	I	1750	O	yes	3.2430	3.2430	0.6037	-0.9257	0.8552	-1.9323	0.2971	-0.3525
8	I	5000	X	yes	3.6990	3.6990	0.8800	-0.1279	0.9756	-0.0247	0.6477	-0.4344
9	I	1750	X	yes	3.2430	3.2430	0.6037	-0.5046	0.8552	-0.1564	0.2971	-1.2138
10	E				-	0.0000	-	-	-	-	-	-
11	E				-	0.0000	-	-	-	-	-	-
12	E				-	0.0000	-	-	-	-	-	-
13	E				-	0.0000	-	-	-	-	-	-
14	E				-	0.0000	-	-	-	-	-	-
15	E				-	0.0000	-	-	-	-	-	-
Nominal Sample size =				8								
Actual number tested =				9								
Dose-averaging estimator				1292.78								
log10 =				3.112								
log-likelihood sums:								-3.3894		-4.8270		-4.6260
likelihoods:								0.0337		0.0080		0.0098
likelihood ratios:										4.2104		3.4436
Individual ratios exceed critical value?				critical=	2.5					TRUE		TRUE
Both ratios exceed critical value?										TRUE		
Calculated maximum likelihood estimate of LD50 =						1329.6						

Final estimate obtained from Maximum Likelihood Calculations

ANNEX 4

CRITERIA FOR CLASSIFICATION OF TEST SUBSTANCES WITH EXPECTED LD50 VALUES EXCEEDING 2000 MG/KG WITHOUT THE NEED FOR TESTING

1. Test substances could be classified in the hazard classification defined by:
2000 mg/kg<LD50<5000 mg/kg (Class 5 in the Globally Harmonised System (GHS)) in the following cases:

a) if reliable evidence is already available that indicates that LD50 to be in the range of class 5 values; or other animal studies or toxic effects in humans indicate a concern for human health of an acute nature.

b) through extrapolation, estimation or measurement of data if assignment to a more hazardous class is not warranted, and

- reliable information is available indicating significant toxic effects in humans, or
- any mortality is observed when tested up to class 4 values by the oral route, or
- where expert judgement confirms significant clinical signs of toxicity, when tested up to class 4 values, except for diarrhoea, piloerection or an ungroomed appearance, or
- where expert judgement confirms reliable information indicating the potential for significant acute effect from the other animal studies.

revcia.
6/06/01

Description of Performance and Confidence Intervals for the Revised Up-and-Down Procedure (UDP) for Acute Oral Toxicity

June 6, 2001

Prepared by:
The UDP Technical Task Force
U.S. Environmental Protection Agency

Submitted to:
The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)

Description of Performance and Confidence Intervals
for the Revised Up-and-Down Procedure for Acute Oral Toxicity

Table of Contents

Title			Page
Executive Summary			1
1.0	Performance and Confidence Intervals for the Revised Up-and-Down Procedure for Acute Oral Toxicity		3
	1.1	Background and History	3
	1.2	Regulatory Applications of Confidence Intervals	3
	1.3	Examples of Regulatory Applications of Confidence Intervals	5
		1.3.1 U.S. Consumer Product Safety Commission	5
		1.3.2 U.S. Environmental Protection Agency (EPA)	6
	1.4	Calculation of Confidence Intervals for the Revised UDP	10
	1.5	Performance Characteristics of the Revised UDP Including Case Examples	11
2.0	LD50 Confidence Bounds for Revised UDP: Statistical Approach and Performance Characterization		17
	2.1	Background	17
	2.2	Confidence Intervals: Definition and Related Terminology	17
	2.3	Classification of Cases and Methods Proposed for Particular Cases	18
	2.4	Confidence Intervals Based on Profile Likelihood	21
	2.5	Simulation Procedures for Measuring the Performance of Confidence Intervals	28
	2.6	Simulation Results	30
3.0	Software		32
	3.1	Purpose and Description	32
	3.2	Quality Assurance/Quality Control	32
4.0	References		34

Appendix A	Performance Characteristics of the Revised UDP Point Estimate and Confidence Interval	35
1.1	LD50 Confidence Bounds for Revised UDP: Technical Specifications and Numerical Programming	35
1.2	Background	36
1.3	Overview of Parametric Approach	37
1,4	Numerical Algorithms for Likelihood Calculations	46
1.5	Classification of Cases	50
1.6	References	53
Appendix B	Tables of Simulation Results	55
Table B-1	Percentages of cases 1-5 among simulated studies	56
Table B-2	Performance measures PM-1-PM-4 (defined in Section 2.5)	60

List of Tables

Table No.	Page
1. Outcomes of the Up-and-Down Procedure: Cases and Confidence Intervals	13
2. Classification of Data Cases for Purposes of Confidence Interval Computation for Case 5	19
3. Example of Case 5	21
4. Example with a Single Partial Response Dose	24
5. Data for Example with Infinite Bounds	26
Appendix A	
A.1 Data for Profile Loglikelihood Example	41
A.2 Example with a Single Partial Response Dose	44
A.3 Classification of Data Cases for Purposes of Confidence Interval Calculation	51
A.4 Classification of Data Cases for Purposes of Confidence Interval Calculation with Computational Procedures	52
Appendix B	
B.1 Percentage of Cases 1-5 among Simulated Studies	56
B.2 Performance Measures PM-1 - PM4 (defined in Section 2.5)	60

List of Figures

Figure No.	**Page**
1. Predicted Percentage of Cases where the LD50 = 1500 mg/ml	15
2. Predicted Percentage of Cases where the LD50 = 3500 mg/ml	16
3. Likelihood Curve for Example 1	22
4. Likelihood Curve for Example 2	23
Appendix A	
A.1. Profile Likelihood Example	42
A.2 . Profile Loglikelihood: Example with Single Partial Response	45

.

List of Abbreviations and Acronyms

AEGL	Acute Exposure Guideline Level
ASTM	American Society for Testing of Materials
ATWG	Acute Toxicity Working Group
CFR	Code of Federal Regulations
CPSC	Consumer Product Safety Commission
CRP	Child Resistant Packaging
EPA	Environmental Protection Agency
FHSA	Federal Hazardous Substances Act
FIFRA	Federal Insecticide, Fungicide and Rodenticide Act
HAZMAT	Hazardous Materials
ICCVAM	Interagency Coordinating Committee on the Validation of Alternative Methods
MLE	Maximum Likelihood Effect
OECD	Organisation for Economic Co-operation and Development
PM	Performance Measure
PPPA	Poison Prevention Packaging Act
QA/QC	Quality Assurance/Quality Control
TG	Test Guideline
UDP	Up-and-Down Procedure

Executive Summary

The draft Revised Up-and-Down Procedure guideline recommends profile likelihood methods, using established theory, for most instances where confidence intervals can be obtained. These are widely used methods that take into account uncertainty in the mean of the population from which the data are drawn. While other types of intervals could have been developed (e.g., bootstrap, isotonic, Bayesian), profile likelihood methods are often used for their practicality and were readily available when the originally proposed Up-and-Down Procedure supplemental test for slope and confidence interval was deleted.

Data gathered under the Revised Up-and-Down Procedure fall into one of five scenarios. Simulations are provided for the performance of the Revised Up-and-Down Procedure in these five cases. Simulations and the fundamental mathematical structure have indicated that in three of these scenarios, standard probit procedures cannot be applied with data generated using the Revised Up-and-Down Procedure. (This can also happen with other multi-treatment-level designs.) Therefore, special statistical procedures are proposed for use in these cases. The point estimates are specified in the test guideline. These circumstances also define availability of the profile likelihood confidence interval and special procedures are proposed for interval estimation.

Calculation of the profile likelihood requires maximizing the likelihood function while holding the term for the LD50 at a fixed assumed value. At each fixed assumed LD50, the likelihood will be maximized by some particular value of the slope. Calculation of the profile likelihood confidence intervals requires calculating the profile likelihood for different values of fixed assumed LD50s with their corresponding profile maximizing slopes and finding the value for which the profile likelihood equals a critical value. This is a computationally-intensive procedure. Consequently, special-purpose software has been developed.

Each of the methods considered can be applied in some scenarios but not in others. In a small percentage of cases no confidence interval would be provided.

1.0 Performance and Confidence Intervals for the Revised Up-and-Down Procedure for Acute Oral Toxicity

1.1 Background and History

Calculation of confidence intervals gives the user a basis for evaluating how to incorporate test results into regulatory applications. Therefore, a confidence interval calculation was included in previous versions of the Up-and-Down Procedure (UDP) guideline (both OECD 1998 and ASTM 1998 and prior). Following deletion of the proposed supplemental procedure from the previous draft Revised UDP, another method was needed to assist the investigator using the UDP to calculate a confidence interval.

The statistical procedure in the previous version of OECD Test Guideline 425 did not produce a true confidence interval because it relied on an assumed value of *sigma* (the slope parameter). This limitation was pointed out in Bruce (1985) and by the ICCVAM UDP Peer Panel (July 2000). While the calculation of the LD50 estimate proposed for the Revised UDP also uses an assumed *sigma*, a separate statistical procedure is proposed for obtaining the confidence intervals for the data. This confidence interval procedure does not rely on the assumed value of *sigma*.

A provision for confidence interval calculation has been added to the statistical analysis of the LD50 estimate from the Up-and- Down Procedure (UDP). Information on the quality of a point estimate and the data from which it is derived are important in understanding the outcome of the test. A confidence interval can be viewed as providing plausible bounds on the value of the LD50 based on the data collected in the particular study. A description of the added feature for calculation of confidence intervals has been inserted at paragraph 40 in the latest revision of the UDP guideline.

An OECD expert group agreed with the addition of the feature for calculation of confidence intervals. Subsequently, the Acute Toxicity Working Group (ATWG) decided to bring the confidence interval insertion to the UDP Peer Panel for comment. Pursuant to these events, a government contract for software development was initiated. The software package for the main test provides (a) information to the experimenter on how many animals are to be dosed and (b) the statistical procedure for estimating the LD50 and confidence interval. A plan for verification of the software package is included in Section 3.0 of this document.

1.2 Regulatory Applications of Confidence Intervals

Statisticians distinguish between point and interval estimation of parameters. Point estimation results in a single value estimate for a parameter, as provided, for example, by the UDP procedure for estimating the LD50. Interval estimation is expressed in a lower and upper bound for an interval that has a known probability of containing the true value of the parameter. That probability is called the confidence coefficient.

To compute a confidence interval, a statistical algorithm needs both the desired confidence coefficient and the experimental data. In the case of the UDP, the experimental data are the doses and responses. The statistical algorithm is designed to compute a 95% confidence interval, which is the typical confidence coefficient in statistical practice. However, the algorithm is not exact but approximate, so that in some situations, the interval will not provide the desired coverage or may provide more than the desired coverage. The results from simulation studies in Appendices A and B of this document will be useful for experimenters to assess if the data and estimated LD50 are producing confidence intervals that are in the same range as simulated intervals that have the desired coverage.

At a given confidence coefficient, the width of the confidence interval is a result of the underlying variability in the dose-response curve. Wider intervals imply less precision in the estimate of the LD50, and also that replications of this experiment with the same compound and animal species under identical conditions could produce meaningfully different LD50 estimates. Moreover, in comparing two different chemical compounds, the widths and locations of the associated confidence intervals provide an indication as to whether the data used to estimate the LD50s lead to estimates precise enough to consider one chemical's LD50 larger or smaller than the other.

Confidence intervals, provided they can be calculated, describe the range of estimates that are consistent with the data seen. In addition, when comparisons of compounds are made using estimated LD50s, confidence intervals give a sense of the robustness of the comparisons. Consequently, any confidence interval is seen as adding descriptively to the data at hand and is not used to exclude information.

Weight-of-evidence deliberations for risk assessments already rely on confidence intervals together with other study details and results. Hazard identification also relies on confidence intervals to assess the meaning of lethality estimates. Such regulatory determinations include:

- decisions about special packaging requirements for products to which children might be exposed,
- registration and reregistration of pesticides,
- review of potential hazard or risk of chemicals to endangered species, and
- hazard identification for consumer and industrial chemicals and mixtures.

Other regulatory instances where confidence intervals are reported include assignment of chemicals or mixtures to toxicity categories used in the regulation of workplace or consumer products, as well as in:

- development of Acute Exposure Guideline Levels (AEGLs; any of three ceiling airborne exposure values for the general public applicable to emergency exposure periods ranging from less than one hour to eight hours);
- routine decisions about child-resistant packaging and labeling;
- classification of substances (e.g., pesticide active ingredients-technical grade);
- for determining hazardous materials (HAZMAT) categories in transport;
- classification of industrial chemicals used in the workplace; and
- classification of mixtures such as pesticide and end-use products (the formulated product).

1.3 Examples of Regulatory Applications of Confidence Intervals

1.3.1 U.S. Consumer Product Safety Commission

Application of Confidence Interval in Evaluation of Hazard and Risk

The confidence interval is important for appropriate evaluation and use of acute toxicity data. An LD50 with a narrow confidence interval that falls within a classification class criteria can be used reliably, whereas an LD50 with a very wide confidence interval (2 mg/kg to 5000 mg/kg) spanning multiple class criteria has to be used very judiciously. The use of numerical values of the LD50 estimate along with the calculated confidence interval becomes more important in a risk assessment (likelihood of injury/illness determination) or when the toxicities of two substances are compared.

The confidence interval is an integral part of a statistical evaluation of toxicity data and its use will be increasingly more important since the number of animals used in testing is being decreased for animal welfare reasons. The number of animals used in a test impacts the size of the confidence interval. Generally, when fewer animals are used, the confidence interval is wider. The width of the confidence interval would determine appropriate use of the data for classification purposes, in risk assessment, or for comparison of toxic potential of two substances, etc.

Regulatory Citations for Acute Toxicity Data including Confidence Intervals:

For a substance to be defined as “hazardous substance”, the Consumer Product Safety Commission under its Federal Hazardous Substances Act (FHSA, 16 CFR 1500.3) requires a two-part determination: 1) that a substance/product has a toxic property, and 2) that it may cause substantial personal injury or substantial illness during or as a proximate result of any customary or reasonably foreseeable handling or use, including reasonably foreseeable ingestion by children. The toxicity data

should be statistically significant and shall be in conformity with good pharmacological practices. A toxicity numerical value such as an LD50 should be accompanied by an index of variability such as a confidence interval.

The Commission also enforces the Poison Prevention Packaging Act (PPPA). The PPPA regulations for exemptions (16 CFR 1700.9 (a)(4)) state:

> "(4) In view of the fact that LD50 values in themselves do not necessarily reflect a true estimate of the overall toxic potential of a substance, LD50 determinations should, where an LD50 value may be calculated, include:
>
> (I) The LD50 value with 95 percent confidence limits; (ii) a slope determination for the dose response curve, including 95 percent confidence limits; and (iii) a description of the statistical method employed in the analysis of such data (with proper citation) as well as the statistical analysis itself."

1.3.2 U.S. Environmental Protection Agency (EPA)

<u>Regulatory Citations for Pesticides under Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA):</u>

40 CFR 158.80 sets forth general policy for acceptability of data as follows:

> "In evaluating experimental design, the Agency will consider whether generally accepted methods were used, sufficient numbers of measurements were made to achieve statistical reliability, and sufficient controls were built into all phases of the experiment. The Agency will evaluate the conduct of each experiment in terms of whether the study was conducted in conformance with the design, good laboratory practices were observed, and results were reproducible."

At 40 CFR 158.202(e)(1) for human health:

> "Determination of acute oral, dermal and inhalation toxicity is usually the initial step in the assessment and evaluation of the toxic potential of a pesticide. These data provide information on health hazards likely to arise soon after, and as a result of short term exposure. Data from acute studies serve as a basis for classification and precautionary labeling."

At 40 CFR 158.202 (h)(2) for nontarget organisms in the environment:

> "The short term acute laboratory studies ... are used to establish acute toxicity levels of the active ingredient to the test organisms; to compare toxicity information with

measured or estimated pesticide residues in the environment in order to assess potential impacts on fish, wildlife and other nontarget organisms; and to indicate whether further laboratory and/or field studies are needed."

Hazard Classification and Risk Assessment of Pesticide Formulations for Human Health:

40 CFR 156.10 provides for hazard labeling of pesticides; Part 152.160 provides for classification of pesticides; and Parts 152.170, 152.171, and 152.175 provide for restricted use of pesticides. Historically, Agency reviewers have tended to consider only the LD50 value in assigning a pesticide formulation to a toxicity category in terms of its oral or dermal toxicity. The traditional acute toxicity study could be relied upon to provide relatively manageable confidence intervals. Confidence limits associated with the LD50 values have generally been reported by the performing laboratories. They are usually included in Agency review summaries.

This situation has changed. With the use of acute toxicity testing protocols that minimize the numbers of animals tested, it becomes more important for Agency toxicologists to consider not only the findings of a study, but also its inherent statistical limitations, in any interpretation and regulatory decision. As a result, in a situation where an LD50 estimate falls so close to a classification boundary that the confidence limits (or bracketing range) include values well below the boundary value, Agency reviewers must take a conservative approach, and classify the test material in the more toxic category. Under these circumstances, the toxicology reviewers would normally feel comfortable with the use of 90% confidence limits, as there would then be only a 5% probability that the LD50 value would be below the lowest value of the confidence interval range. However, they would also have to take into consideration the presence or absence of symptoms of toxicity in the test animals, particularly in situations when severe and/or life-threatening reactions occur at lower dose levels with subsequent recovery and no mortality.

FIFRA Section 25(c)(3) authorizes the Agency to establish Child-Resistant Packaging (CRP) standards, consistent with those under the authority of the Poison Prevention Packaging Act (Public Law 91-601), to protect children from serious injury or illness resulting from accidental ingestion or contact with pesticides. CRP is required for residential use products with an LD50 value of 1500 mg/kg and less, or meeting any of the other toxicity criteria in 40 CFR 157.22(a). If there is a $5% probability that the oral LD50 value is at or lower than 1500 mg/kg, then a toxicology reviewer would recommend the use of CRP. Taking into consideration the emphasis on protecting children from serious injury or illness, an Agency toxicologist would also evaluate the occurrence and severity of toxicological symptoms in an acute oral LD50 study at doses below which mortality occurs.

Environmental Assessment of Pesticides:

Confidence intervals are used in risk assessment for the same purpose as in general statistics to express the "level of confidence" that a sample mean (or other summary statistic) represents the true population mean. Toxicity tests performed for regulatory purposes typically are limited in several ways (i.e.,

sample size, standardized laboratory conditions, etc.). For these reasons, a sample mean (or statistic such as LD50) is generally only a very rough estimate of the actual population being sampled in the test. The confidence interval in this case does describe the level of confidence in the true value, but also serves the reader as a measure of the utility of the test overall. Confidence intervals support compliance with Agency Quality Assurance/Quality Control (QA/QC) principles of precision. Confidence intervals are principally a data QA/QC measure. Point estimates should not to be reported without some measure of precision. Moreover the Agency's QA/QC policies state that the Agency is to use data of known precision. In rating a test result submitted for registration or re-registration of a pesticide, the confidence interval can be considered along with other measures of the validity of the test such as availability of dose response of the test population's tolerance to the pesticide.

Traditionally, toxicity tests for nontarget species are designed to address "dose response" and a narrow confidence interval is an indication of how well a "dose response" was achieved in the study. If the precision of an obtained LD50 study is inadequate, the Agency needs to know that. A good understanding of "dose response" is also useful in risk assessment for extrapolating effects across species and establishing distributional bounds for probabilistic assessments.

The Agency plans to develop methods for probabilistic risk assessments for pesticides which will use confidence intervals from acute tests of nontarget species to describe uncertainty. The uncertainty in the LD50 estimate is an important component in estimating the overall uncertainty in a probabilistic risk assessment. Confidence intervals are necessary for estimating the overall uncertainty/variability in a distribution of risk.

Endangered Species Assessments for Pesticides:

Confidence intervals for the LD50 value are not directly used in assessing effects on endangered species because the intent for endangered species is to protect individuals and not simply the typical representative (i.e., at the population mean). The slope allows the reviewer to determine any mitigation provisions needed to attain an endangered species no-effect level, which is what is necessary under the Endangered Species Act. No-effect levels, such as can be obtained by using the slope in conjunction with the LD50, are used for this purpose. Absent a reliable estimate of the no-effect level, a safety factor is applied to the LD50 value, and the reliability of the LD50 value, as indicated by the confidence intervals is an important feature of the test results.

Setting Acute Exposure Guideline Levels under the Superfund Amendment and Reauthorization Act (SARA):

Acute Exposure Guideline Level-3 (AEGL-3, one of three ceiling airborne exposure values for the general public applicable to emergency exposure periods ranging from less than one hour to eight hours) is the airborne concentration (expressed as ppm and mg/m^3) of a substance at or above which it is predicted that the general population, including "susceptible" but excluding "hypersusceptible" individuals, could experience life-threatening effects or death. Airborne concentrations below AEGL-3

but at or above AEGL-2 represent exposure levels which may cause irreversible or other serious, long-lasting effects or impaired ability to escape.

When a confidence interval is available for an LD50, it may be used to discriminate between studies for use in development of an AEGL-3, to decide whether a study can be used for calculating the LC01 that is the basis for an AEGL-3, or to determine the uncertainty factor in calculation.

U.S. EPA's Policy for Risk Characterization:

The U.S. EPA's Science Policy Council recently issued a Risk Characterization Handbook (EPA 100-B-00-002, Dec. 2000). It focuses on how to integrate "information from the ... components of the risk assessment and [synthesize] an overall conclusion about risk that is complete, informative, and useful for decision makers." Here are some excerpts:

(p. 11) "The overall risk characterization lets the manager, and others, know why the U.S. EPA assessed the risk the way it did in terms of the available data and its analysis, uncertainties, alternative analyses, and the choices made. A good risk characterization will restate the scope of the assessment, express results clearly, articulate major assumptions and uncertainties, identify reasonable alternative interpretations, and separate scientific conclusions from policy judgments."

(p. 13) "Risk characterization communicates the key findings and the strengths and weaknesses of the assessment through a conscious and deliberate transparent effort to bring all the important considerations about risk into an integrated analysis by being clear, consistent and reasonable. Remember, though, unless you actually characterize the assessment, the risk assessment is not complete - - risk characterization is an integral component of every risk assessment. As an example, just giving the quantitative risk estimate ('the number') is not a risk characterization."

(p. 21) "Your specific responsibilities [as a Risk Assessor] are to:

...d) Describe the uncertainties inherent in the risk assessment and the default positions used to address these uncertainties or gaps in the assessment

...f) Put this risk assessment into a context with other similar risks that are available to you and describe how the risk estimated for this stressor, agent or site compares to others regulated by EPA"

(p. 36) "[Elements that affect a Risk Characterization include]:

...f) Variability (Section 3.2.7)

g) Uncertainty (Section 3.2.8)..."

(p. 37) "For each stage of the assessment for human health or ecological risks, the assessor identifies:

a) The studies available and how robust they are (e.g., have the findings been repeated in an independent lab)

b) The major risk estimates calculated, the assumptions and the extrapolations made during the estimated risk calculations, and the residual uncertainties and their impact on the range of plausible risk estimates. Your description of the risk estimate should indicate what you are assessing (e.g., individual, population, ecosystem) and include such things as the high end and central tendency estimates.

...f) Variability (see Section 3.2.7)"

(p. 40) "3.2.7 How Do I Address Variability?

The risk assessor should strive to distinguish between variability and uncertainty to the extent possible (see 3.2.8 for a discussion of uncertainty). Variability arises from true heterogeneity in characteristics such as dose-response differences within a population, or differences in contaminant levels in the environment. The values of some variables used in an assessment change with time and space, or across the population whose exposure is being estimated. Assessments should address the resulting variability in doses received by the target population. Individual exposure, dose, and risk can vary widely in a large population. Central tendency and high end individual risk descriptors capture the variability in exposure lifestyles, and other factors that lead to a distribution of risk across a population."

"3.2.8 How Do I Address Uncertainty?

Uncertainty represents lack of knowledge about factors such as adverse effects of contaminant levels which may be reduced with additional study. Generally, risk assessments carry several categories of uncertainty, and each merits consideration. Measurement uncertainty refers to the usual error that accompanies scientific measurements -- standard statistical techniques can often be used to express measurement uncertainty..."

1.4 Calculation of Confidence Intervals for the Revised UDP

Inserted text at paragraph 40 of the Revised UDP states:

> "40. Following the main test and estimated LD50 calculation, it may be possible to compute interval estimates for the LD50 at specified confidence using a profile-likelihood-based computational procedure. Such an interval utilizes information from the doses where accumulated response was neither 0% nor 100% (intermediate doses). Instead of employing an assumed *sigma*, however, the procedure identifies bounds on LD50 estimates from a ratio of likelihood functions optimized over *sigma* (profile likelihoods). Procedures are also included for certain circumstances where no intermediate doses exist (for instance, when testing has proceeded through a wide range of doses with no reversal or where doses are so widely spaced that each animal

> provides a reversal). Implementing this set of procedures requires specialized computation which is either by use of a dedicated program to be available from OECD or developed following technical details available from OECD."

For many or most studies conducted according to the Revised UDP, standard probit calculations will not be able to provide the basis for a confidence interval. Instead, the Revised UDP uses profile likelihood methods based on established theory for most instances where confidence intervals can be obtained. These are widely used methods that take into account uncertainty in the mean of the population from which the data are drawn. While other types of intervals could have been developed (e.g., bootstrap, isotonic, Bayesian), profile likelihood methods are often used for their practicality and were readily available when the originally proposed UDP supplemental test for slope and confidence interval was deleted.

Profile likelihood confidence intervals are based on the same kinds of functions as the point estimate, namely, the likelihood function and ratios of that function. In addition, the proposed confidence interval uses the same distributional shape assumptions as the point estimate, while making no numeric assumptions about its parameters (i.e., no value for $sigma$ is assumed). In order to reduce such assumptions, this method is computationally intensive using modern methods. Consequently, a specialized program is needed for its implementation. Software will be provided to users on request or through a web site (e.g., OECD's). The OECD Expert Meeting in August 2000 supported this proposal.

The calculation should and does take advantage of established theory, modern computational methods, and previously used and tested algorithms (Rao, 1973; Bickel and Doksum, 1977; Crump and Howe, 1985; Meeker and Escobar, 1995) and utilizes knowledge of the full sample of observations. Results from doses where no or all animals respond does contribute some information on the LD50, even when a point estimate cannot be calculated.

The methodology for this confidence interval has also been used (previously used and tested algorithms) with estimates beside the LD50, including the limit on a benchmark dose (used in U.S. EPA health risk assessments).

Because similar intervals behave well in similar situations, the proposed confidence intervals are expected to perform appropriately for the Revised UDP. The term "behaving well" means that the intervals will have at least the stated coverage probability in simulated trials; that is, at least 95% of simulated '95% CIs' include the true LD50 (see Appendix A).

Just as with the point estimate, there are some circumstances where a standard approach will have computational problems. For example, as outlined in OECD TG 425 paragraph 42 or Revised UDP paragraph 37; there may be only increasing or only decreasing doses throughout the test. Certain solution choices are suggested and included in the special software.

1.5 Performance Characteristics of the Revised UDP Including Case Examples

Five scenarios or cases can be distinguished for the purpose of describing the performance of the Revised UDP as shown in Table 1. Cases 2 and 4 permit estimation of the LD50 and confidence intervals. Cases 1, 3, and 5 do not permit calculation of either an LD50 using the main method, or a confidence interval using the profile likelihood method. Some response patterns for these cases do provide some information about the location of the LD50. More detail on these cases is below.

Case 2 is the standard two parameter probit estimation situation. The case has intermediate response fractions (at least one animal and less than all animals respond) at some dose that is less than a dose where there was no response. Typically, intermediate response fractions will occur at more than one dose. Point estimates and confidence intervals are available.

Case 4 has a single intermediate response fraction occurring between doses that have no response and doses with complete response. The LD50 can be estimated and confidence intervals can be calculated for this case.

Case 1 has three possible response patterns: (a) all animals responded, (b) no animals responded, or (c) the geometric mean dose is lower for animals that responded than for animals that did not respond. Case 1a suggests that the LD50 is likely to be lower than the lowest dose while Case 1b suggests that the LD50 is likely to be greater than the highest dose. Case 1c suggests a reverse dose-response curve, that is fewer responses occur at higher doses. These inferences can be guaranteed to be true, because response is a chance event.

Case 3 has no intermediate response fractions. At some doses, all animals will respond while at lower doses, no animals will respond. This implies that the LD50 is between highest dose with no response and the lowest dose where complete response. Any value between the two doses is a valid estimate for the LD50. No confidence interval can be computed. The situation is likely to emerge from a chemical with a very steep dose-response curve.

There are two possible situations for Case 5. One possibility has an intermediate response fraction at the highest tested dose and no responses at lower doses. This suggests that the LD50 is around the highest tested dose or possibly greater. The second situation has partial response at the lowest tested dose and complete response at higher doses. Here, the LD50 is likely to be at or below the lowest tested dose. For Case 5 data (as for Case 4 data), the LD50 estimate of the software will be the dose with partial response. The confidence interval will be calculated using profile likelihood.

As noted above, data gathered using the Revised UDP fall into one of five types of summary configurations. Simulations and the fundamental mathematics structure have indicated that in three of these configurations, standard probit procedures (e.g., Finney, 1971) cannot be applied with data generated using the Revised UDP. (This can also happen with other multi-treatment-level designs.) Therefore, special statistical procedures are proposed for use in these cases with the Revised UDP.

The point estimates are specified in the Revised UDP. These circumstances also define availability of the profile likelihood confidence interval and special procedures are proposed for interval estimation.

Calculation of the profile likelihood requires maximizing the likelihood (function) while holding the term for the LD50 at a fixed assumed value. At each fixed assumed LD50, the likelihood will be maximized by some particular value of the slope. Calculation of the profile likelihood confidence intervals requires calculating the profile likelihood for different values of fixed assumed LD50s with their corresponding profile maximizing slopes and finding the value for which the profile likelihood equals a critical value. This is a computationally-intensive procedure. Consequently, these will be incorporated into the special-purpose software under development.

Each of the methods considered can be applied in some cases but not in others. In a small percentage of cases, no confidence interval would be provided.

These cases are outlined in Table 1 and Figures 1 and 2.

Table 1. Outcomes of the Up-and-Down Procedure: Cases and Confidence Intervals.

Case #	Definition of Case	Approach Proposed	Possible Findings
1	**No positive dose-response association.** There is no variation in response: a) all animals tested in the study responded, or b) none responded, or c) the geometric mean dose is lower for animals that responded than for animals that did not respond.	No confidence interval proposed, inference related to LD50 questionable.	No statistical results. Possible inferences: 1a) LD50 < lowest dose; 1b) LD50 > highest dose; 1c) reverse dose-response curve
2	**Standard 2-parameter probit estimation.** One or more animals responded at a dose below some other dose where one or more did not respond. The conditions defining Case 1 do not hold. (The definition of Case 2 holds if there are 2 doses with intermediate response fractions, but holds in some other cases as well.)	Profile loglikelihood computations are straightforward.	The LD50 can be estimated and its confidence interval calculated.
3	**No intermediate response fractions.** One or more test doses is associated with 0% response and one or more is associated with 100% response (all of the latter being greater than all of the former), and no test doses are associated with an intermediate response fraction.	Lower bound = highest test dose with 0% response. Upper bound = lowest test dose with 100% response.	High confidence that the true LD50 falls between the two bounding doses. Highest dose with 0% response < LD50 < lowest dose with 100% response.
4	**One partial response fraction, first subcase.** Like Case 3 except that an intermediate response fraction is observed at a single test dose. That dose is greater than doses associated with 0% response and lower than doses associated with 100% response.	Profile loglikelihood calculations to be extended to this case by special computations.	The LD50 can be estimated and its confidence interval calculated.

5	**One partial response fraction, second subcase.** There is a single dose associated with partial response, which is either the highest test dose (with no responses at all other test doses) or the lowest test dose (with 100% response at all other test doses).	Profile loglikelihood calculations to be extended to this case by special computations	The LD50 is estimated and its confidence interval calculated. Possible inference: the LD50 is near the dose with the intermediate response fraction.

Figure 1. Predicted Percentage of Cases - LD50 equal to 1500 mg/kg.

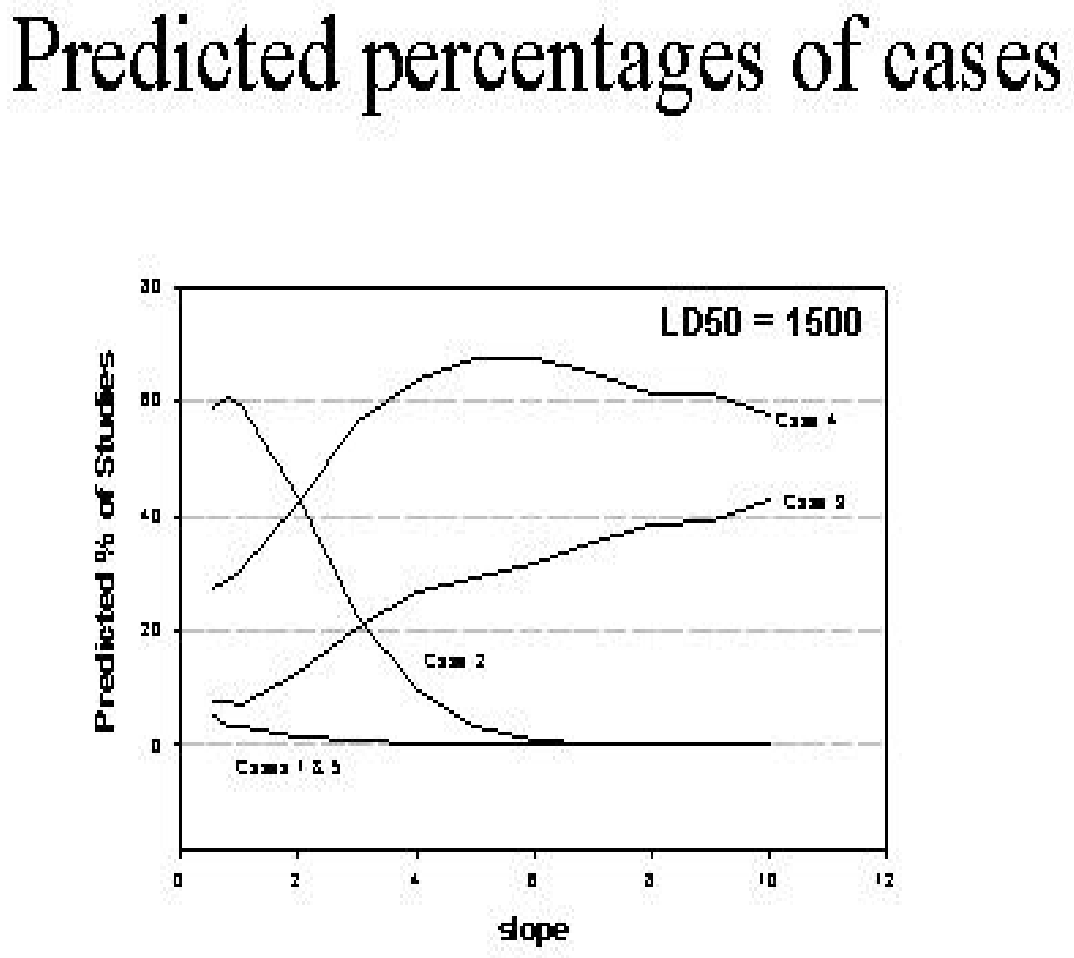
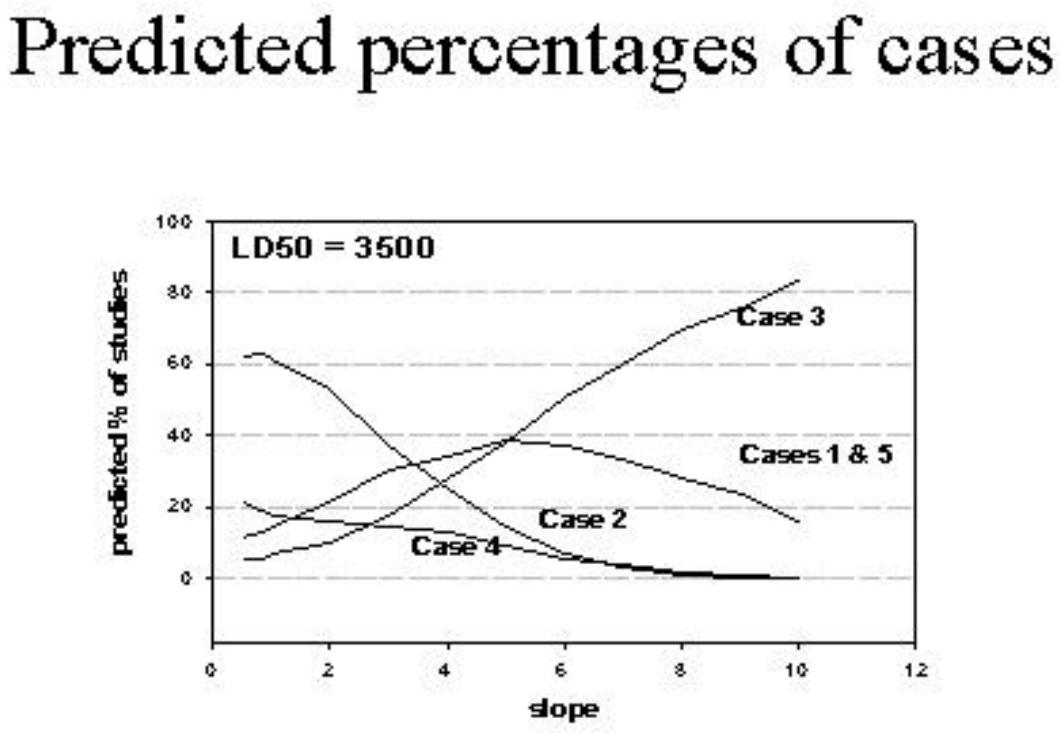


Figure 2. Predicted Percentage of Cases - LD50 equal to 3500 mg/kg.



2.0 LD50 Confidence Bounds for Revised UDP: Statistical Approach and Performance Characterization

2.1 Background.

This section addresses the implementation of confidence bounds for the LD50, for use with acute toxicity data generated in accordance with the Revised UDP. Simulations presented in this document indicate that in a large proportion of cases, standard probit procedures (e.g., Finney, 1971) cannot be applied with data generated using OECD TG 425. Therefore, special statistical procedures are proposed for use with the Up-and-Down Procedure for LD50.

The purpose of this section is to provide an overview of the procedures proposed. Also, simulations are reported to evaluate the performance of the methods proposed. Performance is characterized in terms of the widths of confidence intervals, and in terms of "coverage" probabilities (defined in Section 2.2).

Based on simulations (Section 2.6), it appears that in most cases it will be possible to compute a confidence interval with acceptable performance by one of two methods. In cases where no animals respond at some doses, and all animals respond at some other doses (the latter being greater than the former), the lower bound for the LD50 will be the highest dose associated with no observed responses. Similarly, the upper bound will be the lowest dose associated with response for all animals tested at that dose. In most other cases, it will be possible to compute a bound using the method of profile likelihood (Section 2.4). In particular, it appears that the profile likelihood approach is applicable in most cases where there is only one dose with an intermediate response fraction (neither 0% nor 100% responding), a case that is not handled by standard probit methods. (Proposals for handling various cases are summarized in Section 2.3)

The confidence interval procedures are to be made available in software developed for support of the Revised UDP. The software will also provide point estimates of the LD50 as indicated in the Revised UDP and will evaluate stopping criteria.

The remainder of this section assumes a familiarity with standard probit computations as used in toxicology (Finney, 1971), familiarity with basic statistical procedures (although the definition of a confidence interval is reviewed), and familiarity with the use of Monte Carlo simulation to evaluate the performance of statistical procedures.

2.2 Confidence Intervals: Definition and Related Terminology

Approximate 2-sided 95% confidence intervals will be implemented. Two interpretations of such an interval will be offered in this section. The definition that is most standard is that the probability is 0.95

that the true value of the parameter of interest (here, the LD50) lies within the interval. Here, the parameter of interest is viewed as a fixed constant and the bounds (being based on data) are viewed as random (e.g., Sokal and Rohlf, 1981, particularly Section 7.3). In order for this definition to be useful, the probability of 0.95 must hold at least approximately over the possible values of the parameter of interest, even though the value of that parameter is not know in a given situation.

To understand this interpretation, it may be helpful to reflect on how simulations are used to evaluate a confidence interval (see Section 2.5). In fact, it is common to use simulations to illustrate the concept of a confidence interval (e.g., Sokal and Rohlf, 1981, Figure 7.4).

The probability that the upper and lower bound will enclose the true LD50 is defined to be the *coverage* of the interval. If the coverage of a nominal 95% interval is precisely 95%, then the interval is said to be *exact*. In statistical practice, it is common to use confidence intervals that are not exact but approximate. When intervals are approximate, it is sometimes preferred that they be *conservative*, meaning that the coverage exceeds 95%.

A second interpretation can be particularly helpful for understanding the profile likelihood approach proposed here. According to the second interpretation, a confidence interval for a parameter is to be interpreted as the range of values of the parameter that is consistent with (not excluded by) a particular data set. Thus, Cox and Hinkley state (1974, p. 208) that "foremost is the interpretation that 'such and such parameter values are consistent with the data.' " Confidence intervals can be constructed by inverting statistical hypothesis tests, by defining the confidence interval to be the set of parameter values not rejected using the hypothesis test. In particular, the profile likelihood intervals proposed in this document invert a profile likelihood ratio test.

These two approaches are considered to be consistent. A result given in advanced texts is that a confidence interval with desired coverage can be obtained by inversion of a hypothesis test (e.g., Cox and Hinkley, 1974, Section 7.2; Casella and Berger, 1990, Section 9.2; Bickel and Doksum, 2001, Section 4.2).

2.3 Classification of Cases and Methods Proposed for Particular Cases

Each of the methods considered can be applied in some cases but not in others. In a small percentage of cases, no method of computing a confidence interval is proposed. It is proposed that the selection of a method be based on the classification of cases displayed in Table 2. (Development of this scheme has benefitted from discussions with the OECD acute avian statistics group. See Table 2 footnote.) The rationale for the decisions indicated in this table is as follows.

Case 1. With the stopping rules indicated for the Revised UDP, this case appears to be possible only if testing is stopped at a limit dose (based on non-response for three animals tested in sequence at the dose). No methods are proposed here for cases where there is not an observable relationship between dose and response. In some cases, a binomial test may be used to establish that the LD50 is above or

below the range of doses tested, but a significant binomial test requires testing of five or more animals at the same dose, and binomial tests use only data from a single test dose. Some procedures that may be applicable in this case have been developed for avian

Table 2. Classification of Data Cases for Purposes of Confidence Interval Computation for Case 5

Case #	Definition of Case	Approach Proposed
1	**No positive dose-response association.** There is no variation in response (all animals tested in the study responded, or none responded), or the geometric mean dose is lower for animals responding than for animals not responding	no confidence interval proposed, inference related to LD50 questionable.
2	**Standard 2-parameter probit estimation.** One or more animals responded at a dose below some other dose where one or more animals did not respond. The conditions defining Case 1 do not hold. (The definition holds if there are two doses with intermediate response fractions, but holds in some other cases as well.)	profile loglikelihood computations are straightforward
3	**No intermediate response fractions.** One or more test doses is associated with 0% responses and one or more test doses is associated with 100% responses (all of the latter being greater than all of the former), and no test doses are associated with an intermediate response fraction.	lower bound = highest test dose with 0% response. upper bound = lowest test dose with 100% responses.
4	**One partial response fraction, first subcase.** Like Case 3, except that an intermediate response fraction is observed at a single test dose. That dose is greater than doses associated with 0% responses and lower than doses associated with 100% responses.	profile loglikelihood calculations to be extended to this case by special computations
5	**One partial response fraction, second subcase.** There is a single dose associated with partial response, which is either the highest test dose (with no responses at all other test doses) or the lowest test dose (with 100% responses at all other test doses).	profile loglikelihood calculations to be extended to this case by special computations

Acknowledgement. The development of this scheme was helped by conversations with the OECD avian acute statistics group, which has developed a similar classification (report in press). The avian scheme differs in some points.

acute testing (report in press).

Case 2. In cases where standard probit computations can be applied, it appears that application of the profile likelihood (described in Section 2.4) will be straightforward. The profile likelihood approach is already used in this situation in the U.S. EPA benchmark dose software.

It is common to require, as a condition for probit analysis, that there are at least two test doses with partial response fractions (response fractions not 0% and not 100%). Case 2 as defined here includes all the cases with at least two partial response fractions, but includes other cases as well. In the definition of Case 2, one or more animals respond at some dose, such that one or more do not respond at some higher dose (Silvapulle, 1981).

In addition, the geometric mean dose must be higher for animals that respond than for animals that do not respond. The second condition is indicated in Revised UDP as a requirement for inferences regarding the LD50.

In standard probit analysis, bounds of a confidence interval may be infinite. The standard approach for detecting whether the bounds are infinite is based on a test of the statistical significance of the slope parameter. An analogous procedure can be used with the profile likelihood approach.

Case 3. When there are no partial response fractions (along with other requirements of the case, as indicated in Table 2), for technical reasons the profile loglikelihood approach apparently cannot be applied in a straightforward manner. In this case, it seems that any dose within the interval bounded by the highest dose with no responses, and the lowest dose with 100% responses, would be about equally valid as an estimate of the LD50. It seems natural to consider whether those two doses can function in practice as an approximate confidence interval, and there does not appear to be any alternative for defining bounds in this case.

For Case 3, the proposed bounds are not designed to achieve a specific confidence level. Rather, the approach is to ask what is the realized confidence level, if bounds are computed in a certain way.

Case 4. When there is a single partial response (along with other requirements for the case, as indicated in Table 2), the profile loglikelihood can be applied using special computations developed by the ICCVAM Acute Toxicity Working Group. Some technical details are given in Appendix A..

Case 5. This is an infrequent case, which appears to occur primarily when an LD50 is close to a bound. Table 3 is an example of Case 5, generated in a simulation of the Revised UDP.

Table 3. Example of Case 5.

dose (mg/kg)	number tested	number responding
1.0	6	5
1.5	2	2

In the simulations, test doses are restricted to the range 1-5000 mg/kg. For the result displayed in Table 3, testing was probably stopped when three animals tested in sequence at 1 mg/kg all responded.

It could be concluded that the LD50 is more than likely to be below 1.5 mg/kg. A profile likelihood calculation can be done.

2.4. Confidence Intervals Based on Profile Likelihood

This section provides a non-mathematical overview of profile likelihood computations proposed for use when the data from a given study is assigned to Case 2 or Case 4. The methods are illustrated using hypothetical data sets, which were generated in simulations of the Revised UDP.
Some technical details and formulae are provided in Appendix A. The material in this section is not needed in order to understand the evaluation of performance of the methods using simulations, which is found in the sections that follow. However, it is desirable to understand the following points: First, the type of bounds proposed will be infinite in some cases. More precisely, both the upper bound and lower bound will be finite or both bounds will be infinite. This is as in standard probit analysis. Second, the methods proposed cannot be implemented by plugging data into a formula. Specialized computing skills such as numerical optimization are required for implementation. For the numerical aspects, there are multiple alternative algorithms that may be used without actually changing the statistical approach.

Explicit descriptions of the profile likelihood approach are found in Barndorff-Nielsen (1991), Davidson and MacKinnon (1993), and Meeker and Escobar (1995), among other sources. Implicit justification for the approach is found in any theoretical statistics book if it is noted that (I) confidence intervals can be constructed by inverting statistical tests (Section 2.2) and (ii) the method proposed inverts a likelihood ratio test that is ordinarily presented. (These references are somewhat technical. The point here is to confirm that the general type of approach suggested is well established in statistics.) The method has been widely used in connection with nonlinear statistical models, and descriptions can be found in literature associated with various applications. Barndorff-Nielsen (1991) uses the term *profile likelihood* to denote the particular variant of a likelihood function that is used here, while other authors do not specifically name that variant. Barndorff-Nielsen (1991) also reviews refinements of the approach.

According to the approach proposed, statistical results are based on likelihood curves. Figures 3 and 4 provide two examples of likelihood curves, based on hypothetical data examples. Formulae for the likelihood curves are provided in Appendix A. Points to be emphasized

Figure 3. Likelihood Curve for Example 1.

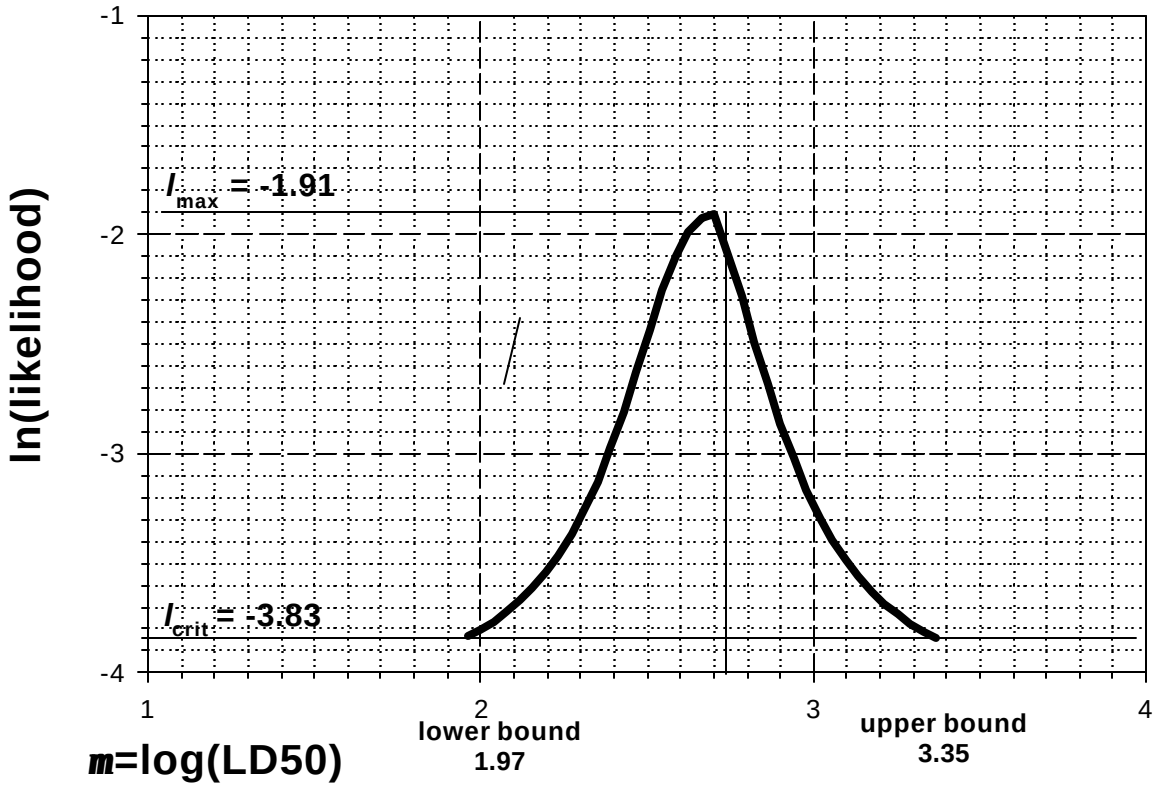
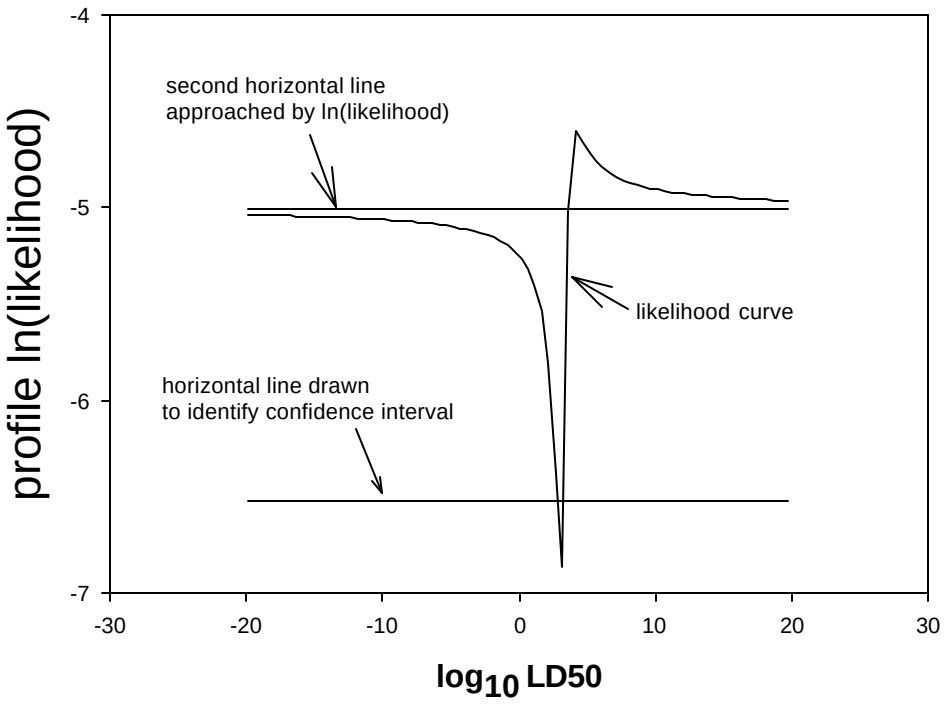


Figure 4. Likelihood curve for Example 2.



include that each distinct data set is associated with a distinct likelihood curve, from which can be read the statistical results (confidence bounds as well as a point estimate) for that data set. The likelihood curve also depends on the type of dose-response function that has been assumed. Revised UDP specifies the use of probit models. A logit model would also have an LD50 and a slope, closely analogous to the probit LD50 and probit slope, but for a given data set the likelihood curves for a probit model and a logit model would not be identical.

For present purposes, it is helpful to think of the likelihood curve as providing levels of relative support that a specific data set give to different choices of parameter values (Edwards, 1992). In particular, the LD50 value with highest likelihood is the *maximum likelihood estimate* (MLE) and may be considered for a point estimate. (However, Revised UDP specifies that the LD50 point estimate will be a MLE based on an assumed slope.) It turns out that standard probit calculations generate maximum likelihood estimates although the likelihood is not computed explicitly (Finney, 1971).

If this notion of likelihood-as-support to calculation of confidence bounds is extended, it seems that values inside the confidence interval should have higher likelihood than values outside the interval. The upper and lower bounds for the confidence interval, it seems, should have equal likelihood (see Figure 3). This notion is the basis of the graphical approach described with the following examples.

Example 1. The following data were generated in a simulation of Revised UDP..

Table 4. Example with a Single Partial Response Dose.

dose (mg/kg)	number tested	number responding
175.0	2	0
553.4	3	2
1750.0	1	1

likelihood values (see text for explanation)
maximized loglikelihood = -1.910
loglikelihood for bounds =-3.830= -1.91 - 1.92
95% CI for the LD50 = 93 - 2258 mg/kg based on method of profile likelihood

Here there is only a single partial-response dose and so standard probit programs cannot be used to generate an estimate of the LD50. The likelihood curve associated with these data is displayed in Figure 3 [the natural log of the likelihood is graphed. Use of ln(likelihood) is conventional in statistics for computations with likelihoods.]

The confidence bounds can be computed graphically using Figure 3, by the following steps:

(1) There are two parameters in the probit model, namely the slope and LD50, but the curve displayed is a function of the LD50 only. A 2-parameter likelihood can be defined which can be graphed in three

dimensions. In the context of Revised UDP, the LD50 is of primary interest. In this context, the slope is said to be a nuisance parameter. Therefore, it does seem useful to obtain a likelihood curve for the LD50 alone, if that is possible.

One way to eliminate the slope, as used in Revised UDP point estimation and stopping rules, would have been to assume a value for the slope. Here, a more computationally intensive approach has been used. The approach proposed is the detail that defines the profile likelihood approach specifically, as a type of likelihood approach. According to the profile likelihood approach, at each value of the LD50 the slope value is used that maximizes the 2-parameter likelihood.

Since the profile likelihood curve is the only likelihood curve that will be used in this document, the profile likelihood (for the LD50 eliminating the slope) will be referred to as "the likelihood curve" although, to be more exact, it should be referred to as the "profile likelihood curve."

(2) For the hypothetical data, the likelihood function has a peak where the log(LD50) has the value of approximately 2.7 (i.e., at an LD50 value of 553 mg/kg). Note that the value of 553.4 mg/kg is the middle dose in this example, the dose with an intermediate response fraction. This value would not be a bad choice of a point estimate for the LD50 for these data.

(3) The subsequent computations require the peak value of the ln(likelihood). In this particular example, special computations are needed to get the maximized (peak) ln(likelihood), which are presented in Appendix A. For the data considered here, these computations yield a value of -1.91 for the maximized ln(likelihood), which is evidently consistent with the curve in Figure 3. In cases where standard probit calculations can be applied, computation of the maximized ln(likelihood) involves a different procedure, as in Example 2 below.

(4) An approximate lower bound for the LD50 can be read from the likelihood curve as follows. A horizontal line is drawn at a (log) likelihood value of -3.83, a value which is computed with a formula below. Referring to Figure 3, this line is seen to intersect the likelihood curve to the left of the curve peak, at an LD50 value of 92 mg/kg ($\log_{10}$(92)=1.965). Therefore, the value of 92 mg/kg is taken to be the lower bound for the LD50.

A similar approach is used for determining the upper bound of 2258 mg/kg ($\log_{10}$ = 3.35). The upper bound value is the dose value where the horizontal line crosses the likelihood at a second point, to the right of the point estimate.

The Y-axis value of the horizontal line (-3.83 for this example) is calculated with the following formula which has been developed by mathematical statisticians:

$$\text{ln(likelihood) for bound} = \text{maximized ln(likelihood)} - 1.92$$

For the example, a maximized ln(likelihood) value of -1.91 has been calculated, so the Y-axis value for

the horizontal line is -1.91 - 1.92 = -3.83.

In the formula above, the value of 1.92 is appropriate for computation of a 2-sided 95% interval. A different value would be used to compute a 90% interval, and so on. Technically, the value to be used is taken from tables (or the electronic equivalent) of a chi-square distribution with one degree of freedom.

To see why these computations make sense, reflect again on the notion that the likelihood is a measure of relative support that the data give to alternative choices of an LD50. The graphical approach separates the possible choices of an LD50 into two sets based on their likelihoods: The confidence interval comprises LD50 candidates with ln(likelihood) above the horizontal line, while LD50 candidates outside the confidence interval have ln(likelihood) below the horizontal line. The two bounds are dose values with equal likelihood. The procedure seems natural if LD50 candidates with higher likelihood are regarded as better supported by the data.

(5) Reflection on the procedure just described indicates a possible problem. The likelihood curve was graphed over a finite range. The graphical approach assumes that the ln(likelihood) remains below the horizontal line for LD50 values not graphed. If not, then the bounds are infinite. However, as mentioned previously, there is a way to determine if the bounds are finite or infinite. Use of the formula in this case indicates that the bounds are finite.

Example 2. The following hypothetical data were also generated in a simulation of Revised UDP.

Table 5. Data for Example with Infinite Bounds

dose (mg/kg)	number tested	number responding
175.00	1	0
553.40	2	0
1750.0	3	1
5000.0	4	1

Probit results: slope = 1.02, estimated LD50=14223 mg/kg

Standard probit calculations (Finney, 1971) can be performed in this case. Probit results for the LD50 and slope are displayed in a table footnote. According to standard probit calculations, the bounds for the LD50 are infinite in this case.

The likelihood curve based on these data is displayed in Figure 4.

The curve can be used for the point estimation because the likelihood curve has an unambiguous peak. If the graph is plotted over a more narrow range than that used for Figure 4, it can be seen that the peak actually does correspond to the probit LD50 estimate. In fact, standard probit calculations do generate the maximum likelihood estimates (Finney, 1971).

Next we need the maximized ln(likelihood). In this case, the computations are different from those used for Example 1. When the standard probit calculations apply (as in this example but not in Example 1), the maximized ln(likelihood) is computed by plugging the probit estimates of the slope and LD50 into the two-parameter likelihood formula. The two-parameter likelihood formula is given in Appendix A.

As in Example 1, a horizontal line can be drawn separated from the peak ln(likelihood) by a value of 1.92 units in the direction of the Y axis. The result of this step is the lower of two horizontal lines drawn on the graph (see Figure 4). In this case, although the likelihood curve dips below the horizontal line, the set of dose values with ln(likelihood) above the line (those values not excluded based on our data) stretches to infinity in each direction. Consistent with the results of standard probit computations for this case, the profile likelihood confidence bounds are considered infinite.

Note that if the likelihood curve had been viewed over a narrow range of LD50 values around the peak, one might have concluded that the upper bound was probably infinite, but might be misled to suppose that the lower bound is finite. This problem can be resolved as follows. Observe that in this case as the LD50 approaches infinity in either direction, the likelihood curve approaches a second horizontal line (refer to Figure 4). In fact, it appears that in all cases the likelihood curve will approach some line in this way, and the location of that line can be determined. (The formula is provided in Appendix A.) Evidently, the bounds are finite if and only if the second line is located below the first in the Y-axis direction.

Computer algorithms, particularly handling of infinite bounds. Despite what these examples may suggest, it is not proposed that in practice the bounds will be obtained by literally drawing lines on graphs. A computer program will be used to perform analogous computations. However, understanding of the graphical approach just given can provide an appreciation of the types of computer algorithms required to implement the approach. Three types of specialized computer algorithms are evidently needed.

(I) The approach requires that we compute the maximized value of the ln(likelihood). When the results of a study fall in Case 2, an optimization (peak finding) algorithm is required. Standard probit calculations (Finney, 1971, Ch. 2) represent an appropriate optimization algorithm in this case, and that approach has been used in simulations reported in the following sections.

(ii) Computation of the bounds requires us to identify values of the LD50 that have specific values of the ln(likelihood). For the simulations reported in this document, a bisection algorithm has been used.

(iii) In Example 1, it was explained how the slope is eliminated from the likelihood function when using the profile likelihood method. (For a given value of the LD50, use the slope value that maximizes the likelihood.) Consequently, another optimization routine is needed. In simulations, a type of weighted Gauss-Newton algorithm, also termed a scoring algorithm, has been used. This is a type of optimization method widely used in situations such as probit fitting (Nelder and Wedderburn, 1989).

Each of these three operations involves a kind of iterative search procedure, meaning that some kind of initial guess is developed for a quantity to be computed and that guess is refined in an iterative fashion, until further refinements seem to have no practical effect. The implementation of these types of algorithms requires a specialized type of computing skill.

For each of the three operations identified there are various algorithms that may work. The choice of an algorithm is not fundamental to the statistical method, but can affect the performance of a computer program in some ways. If a relatively poor algorithm does not produce incorrect results, computing speed may be slowed, or the algorithm may occasionally fail to produce results because of a variety of numerical phenomena.

2.5. Simulation Procedures for Measuring the Performance of Confidence Intervals.

In previous work, we have used simulations to evaluate the performance of OECD TG 425 for the purpose of estimating the LD50. In these simulations, values were assumed for the slope, LD50, and starting dose, and numerous data sets were simulated. In that situation, estimates of the LD50 close to the true value are considered desirable. Therefore, performance could be evaluated by considering the percent of simulated studies yielding LD50 estimates in some sense close to the true value, say within some factor of the true value.

Analogous simulation procedures have been used here to evaluate the performance of the proposed confidence intervals. As with previous simulations, values are assumed for the LD50, the slope, and the initial starting dose. For a given combination of assumed values of these parameters, we simulate a large number of studies. The simulation results are used to compute measures of performance. While the procedure for simulating data sets is identical to the procedure used in evaluation of point estimates, different performance indices are computed from the simulated data.

To assess the performance of the confidence intervals, we report four measures of performance, which are denoted PM1, PM2, PM3, and PM4 in the tables of simulation results.

PM1. This is the estimated percent of studies that have finite confidence bounds. (The bounds are both finite or both infinite.) It is desirable to have narrow confidence bounds, but it is not clear that the occurrence of very wide bounds should be viewed as a drawback for the method of computing confidence bounds, versus as a drawback of the study design. In any case, the index seems to provide useful information.

PM2. This is the coverage, which is the fraction of studies for which the true LD50 falls inside the confidence interval (above the lower bound and below the upper bound). For each of, say, 1000 simulated studies, the confidence intervals are computed with the procedures proposed, and the study is scored as either enclosing the true LD50 or otherwise. PM2 is then the percent of the 1000 simulated studies with bounds that enclose the true LD50. In cases where the bounds were infinite, they were scored as enclosing the true LD50.

By the definition of a 95% confidence interval, the ideal value for PM2 is 95%. Ideally, PM2 will not vary when the LD50, slope, and initial test dose are varied.

PM3. PM3 and PM4 are alternative measures of the typical widths of confidence intervals. PM3 is the median ratio of the upper bound to lower bound. (The ratio upper/lower is computed separately for each of, say, 1000 simulated studies. PM3 is then the median of the 1000 ratios.)

In cases where the bounds were infinite, the ratio was coded as greater than 1000 (>1000). Ratios that were finite but greater than 1000 were also coded as simply >1000. (Otherwise some confidence intervals with finite bounds would be coded as more narrow than intervals with infinite bounds.) For many of the situations where a slope of 0.5 or 0.8 was simulated, over 50% of simulated studies had infinite slopes (that is, PM1>50%). (See Table B.2 of Appendix B.) In these cases , the value of PM3 is >1000. In a few cases, PM3 was >1000 when PM1 is slightly below 50%, because of some intervals that are very wide but not infinite.

Use of a value of 1000 is somewhat arbitrary but this choice does not effect the median ratio unless the ratio exceeds 1000 for at least 50% of simulated studies. We suggest that if the median ratio is greater than 1000, there is not much practical value in quantifying the proportion of confidence intervals with infinite bounds, versus with bounds that are finite but separated by a factor of 1000 or more.

In several cases where a steep slope is assumed, PM3 is equal to 3.2. This is the ratio of adjacent test doses, except in some cases where a test dose is a limit dose. In these cases, many data sets fall under Case 3, for which all doses are associated with either 0% response or 100% responses..

PM4. This is a second measure of typical width, the median standardized width of the confidence interval. For each simulated study (say for 1000 simulated studies), we compute the quantity:

$$\text{standardized width of confidence interval} = 100 * \frac{\text{upper bound - lower bound}}{\text{true LD50}}$$

This ratio is computed for each of, say, 1000 simulated studies. PM4 is then the median of the 1000 standardized widths.

In cases where the bounds were infinite, the standardized width was scored as >100,000. This is comparable to use of a code of >1000 for PM3 given the approximate relationship between the two indices.

In interpreting these measures, it may be useful to consider the coverage measure PM2 jointly with measures of width (PM3 or PM4). If the coverage is larger than 95% and the intervals appear undesirably wide, then a case can be made for refining the statistical procedure to yield more narrow bounds, with coverage closer to the ideal value of 95%.

OECD standard simulation scenarios for acute mammalian guidelines. Simulations have been conducted based on two sets of scenarios. (the term scenario is used to mean a combination of true LD50, true slope, and initial test dose.)

The first set of scenarios comprised 45 combinations of slope and LD50, with the initial test dose set to 175 mg/kg in each case. The value of 175 mg/kg is the Revised UDP default initial test dose, to be used when there is no reliable information to indicate a better initial test dose. The combinations of slope and LD50 for this set are the same as for the second set.

The second set of scenarios comprises 112 combinations of slope, LD50, and initial test doses. This set of scenarios has been developed by OECD workgroups for evaluation and comparison of acute toxicity designs. For this set, initial test doses were initially specified in terms of percentiles of the tolerance distribution. The test doses were then calculated from the slopes and LD50s. In simulations of the Revised UDP, test doses are restricted to the range of 1 to 5000 mg/kg. Therefore, combinations with an initial test dose outside that range have been deleted.

In this set, scenario number 95 has been modified for the simulations reported here, by changing the initial test dose from 4870 to 4750 mg/kg. The LD50 is 3000 mg/kg for this scenario so that testing tended to be concentrated on the two doses 4870 and 5000 mg/kg, the latter being the limit dose. The original value of 4870 mg/kg is unrealistically close to the limit dose of 5000 mg/kg and the scenario was unmanageable numerically because of a large number of numerical overflows. When the initial test dose was changed to 4750 mg/kg, no further difficulties were encountered. (No numerical problems were encountered with any of the other scenarios, after some refinements of the algorithms.)

Additional details of simulation. The performance measures PM1-PM4 were computed only using data for Cases 2-4, because it is only for those cases that statistical methods are proposed in this document. For example, PM1 is then the percent of studies in Cases 2-4 that have finite intervals. However, the percentages of studies assigned to different cases were computed using the data for all cases.

For each scenario, a minimum of 1000 studies was simulated. Because confidence intervals were computed for Cases 2, 3, and 4, the combined number of simulated studies for those 3 cases was fixed at 1000 for each scenario while the total number simulated studies per scenario was variable but always greater than 1000.

As in previous simulations, the range of test doses has been restricted to the range of 1 to 5000 mg/kg.

2.6 Simulation Results

Two types of simulation results are provided in Appendix B.

Table B.1 of Appendix B provides percentages of Cases 1-5 for each scenario. (See Table 2 of this

Section for the definitions of these cases.) A combined percentage is reported for Cases 1 and 5. Table B.1 contains the results for both sets of scenarios, those with the initial test dose fixed at 175 mg/kg and those with initial test dose varied.

The case frequencies are informative regarding how often particular procedures can be applied. In particular, the low frequency of Case 2 in many scenarios supports our assertion that standard procedures of probit analysis will often not be applicable with TG 425. Cases 1 and 5 occur with relatively high frequency when the true LD50 is close to a limit dose. This is probably a consequence of instances where a particular stopping rule is invoked, namely that testing is stopped if three animals tested in sequence at 5000 mg/kg do not respond, or if three tested in sequence at 1 mg/kg all respond.

The relative frequencies of different cases depends strongly on the slope, for obvious reasons. If the slope is steep, then the percentages of animals responding changes from 0% to 100% within a narrow range of dose values, and the possibility for obtaining a partial response percentage therefore relatively small.

Table B.2 provides the values of performance measures PM1-PM4 (defined in Section 2.5) for each Scenario. Overall, the results seem to suggest acceptable performance of the methods proposed.

The results indicate a strong dependence on the slope. As the slope increase, the percentage of infinite bounds is lower (PM1), the coverage increase (PM2), and the intervals become more narrow (PM3, PM4).

With regard to coverage (PM2) the ideal value is 95%, and ideally the coverage will not depend on the slope. Therefore, the PM2 values of 99%-100%, associated with steep slopes, are not necessarily to be viewed favorably. However, in the steep-slope situations, the confidence intervals tend to be narrow (PM3, PM4). Thus, the conservatism of the methods when the slope is steep (as quantified by PM2) do not seem to represent a serious drawback of the methods proposed.

3.0 Software

3.1 Purpose and Description

Because the Revised UDP is relatively complex statistically, dedicated software has been developed to integrate all statistical features of the test, including a) multiple stopping criteria; b) estimation of an LD50; and c) provision of confidence intervals , together with their appropriate places in the laboratory protocol. This software was developed for a Windows environment and is accompanied by a user manual. The software and manual are designed to be readily understood and implemented by scientists outside the U.S. who may have limited facilities and English comprehension. It will be a stand-alone package designed for analysis only, with provision for an investigator to create reports that include animal identifiers that match those in a laboratory's standard data maintenance files, thereby permitting data verification.

Development of this software is being carried out under contract to the U.S. EPA, through work assignments 4-06 and 5-03 of Contract No. 68-W7-00285. Building the package follows practice for verification, which is an abbreviated form of standard practice such as that outlined by the FDA draft guidance for industry on general principles for software validation. The FDA guidance states:

> "Verification is defined in 21 CFR 820.3(aa) as "confirmation by examination and provision of objective evidence that specified requirements have been fulfilled." In a software development environment, software verification is confirmation that the output of a particular phase of development meets all of the input requirements for that phase. Software testing is one of several verification activities, intended to confirm that software development output meets its input requirements. Other verification activities include walkthroughs, various static and dynamic analyses, code and document inspections, both informal and formal (design) reviews and other techniques".

The model of verification is not unlike the QA/QC Check of the Benchmark Dose System (BMDS) Software for the U.S. EPA (Contract No. 68-C9-8007, Work Assignment 1-10, December 1999).

Completion of all construction, testing, and documentation is scheduled for summer 2001.

3.2 Quality Assurance/Quality Control

Software requirements are being set out by the U.S. EPA and the contractor regarding environment, input/output/functions, user interfaces, error handling; design is considering implementation (coding) issues; and testing will be performed to ascertain that the package does what it is designed to do. Some of this testing will be in the form of stressing the program by pushing it to unusual circumstances (and sample data sets are currently under construction). Some of these data sets generally can be described by the case descriptions in section 2 of this document. The sets specifically encompass, however, such situations as possible data entry errors and the various stopping circumstances, as well

as unusual dose magnitudes. Some of it will constitute simulations characterizing the behavior of Revised UDP that can be compared to independently programmed output regarding Revised UDP behavior. When completed, these activities will constitute a verification of the analysis package.

At the first stage, an outline of the program has been created, identifying its structure (with data, calculation, and report modules, and, for testing, a simulation module), how modules will interact, what each module will do and, as appropriate, the mathematics for those operations; enumerating the possible configurations of data and which will and will not give numeric solutions; describing messages (prompts, warning, error) from package to user and their circumstances; and outlining the testing and simulation processes. Concurrently, an outline of the user manual was delivered.

4.0 References

Barndorff-Nielsen, O.E. 1991. Likelihood theory. Chapter 10 in D.V. Hinkley, N. Reid, and E.J. Snell (eds) *Statistical Theory and Modelling*. Chapman and Hall.

Bickel,P.J. and K. A. Doksum. 2001. *Mathematical Statistics: basic ideas and selected topics*. Volume 1. (2nd ed.) Prentice Hall.

Casella, G., and R.L. Berger. 1990. *Statistical Inference*. Duxbury.

Edwards, A.W.F. 1992. *Likelihood*. 2nd ed. Johns Hopkins Univ. Press.

Davidson R., and MacKinnon, J.G. *Estimation and Inference in Econometrics*. Oxford U. Press.

Finney, D.J. 1971. *Probit Analysis*. 3rd ed. Cambridge U. Press.

McCullagh, P., and J.A. Nelder. 1989. *Generalized linear models*. (Second ed.) Chapman & Hall/CRC.

Meeker, W.Q., and Escobar, L.A. 1995. Teaching about confidence regions based on maximum likelihood estimation. The Amer. Statistician. 49(1):48-52.

Silvapulle, M.J. 1981. On the existence of a maximum likelihood estimators for the binomial response model. J. Royal Statist. Soc. Series B 43(3):310-313.

Appendix A

Performance Characteristics of the Revised UDP Point Estimate and Confidence Interval

1.1 LD50 Confidence Bounds for Revised UDP: Technical Specifications and Numerical Programming

This appendix provides technical detail and mathematical formulas, and supports technical peer review and programming.

The preliminary approach, described in this Appendix, was to limit the numerical search for a bound to a finite interval above or below a point estimate of the LD50. This approach was used because no procedure was readily available to determine from the data, *a priori*, whether the bound is finite or infinite. It was suggested that the search interval may be made sufficiently wide so that, if a bound is outside the interval, it might be considered infinite for practical purposes.

However, it appears that there is actually a criterion that can be used to determine whether the bounds are finite or infinite. The probit model can be parameterized in terms of $\mu = \log_{10}(\text{LD50})$ and the slope ($\$$). According to the method of profile likelihood, the decision of whether a value of μ is inside or outside the confidence interval is made by optimizing the slope parameter with μ fixed at the value of interest, and thus obtaining an optimized loglikelihood value corresponding to a particular μ value. The value of μ in question falls within the confidence region if and only if the maximized loglikelihood is greater than or equal to a critical loglikelihood that can be denoted as l_{crit}. The computation of l_{crit} is as described in this Appendix.

As the value of μ is taken toward infinity in either direction, and the slope is optimized for each value of μ, the optimized slope value is observed to converge to zero. The loglikelihood is observed to converge to a value that can be computed directly, by substituting for each predicted response percentage the pooled response percentage $p_{\text{pooled}} = \sum_{i=1}^{g} r_i \Big/ \sum_{i=1}^{g} n_i$ where r_{i} and n_{i} are the numbers of animals that respond and the number tested at the ith of g dose levels tested.

This behavior can be understood as follows. For definiteness, consider computation of the lower bound. As μ is taken toward negative infinity, the value of $\$$ approaches zero. If $\$$ did not approach zero, then all of the predicted response probabilities would go to zero. However, the methods are applied only when some animals respond and others do not. Therefore, $\$$ goes to zero to fit a mixture of animals that responded and did not respond as μ is taken to infinity. When $\$$ is close to zero, the doses become, for purposes of probit analysis, about the same dose. (For purposes of probit analysis, the magnitude of dose ratios is considered relative to the slope.) Therefore, as μ is taken to infinity and $\$$ optimized at each value of μ, the fitted probit line approaches a line connecting the point (μ,probit(0.5)) to the point ($\bar{x}$,probit(p_{pooled})) where $\bar{x}$ is the mean log dose.

Therefore, the criterion for determining whether or not the bounds are finite is as follows. Let l_{pool} denote the value of the loglikelihood computed with each response percentage set equal to p_{pooled}. Since the profile loglikelihood will approach l_{pool} as μ approaches ± 4, the bounds are finite if and only if l_{pool} is less than l_{crit}.

Figure 3 in Section 2.4 of this document, and the associated discussion of Example 1 is misleading. If the loglikelihood is graphed over a sufficiently wide range of doses, the loglikelihood is seen not to be convex and the nonlinear equations that define the bounds have more than two roots. (In the graphs of Section 2.4 of this document, the curve crosses the line more than twice.) In this case, according to the criterion just described, the lower bound as well as the upper bound is infinite, which is also the result obtained with the standard probit methods.

1.2 Background. The ATWG proposes to implement confidence bounds for the LD50, for use with acute toxicity data generated in accordance with the Revised UDP. The method for calculating the confidence interval will be available in software developed to support the Revised UDP; this software will also provide point estimates of the LD50 and will evaluate stopping criteria. The decision to develop new confidence interval procedures is based on simulations that indicate that standard procedures (for analysis of data under a 2-parameter probit model) will very often not be applicable with data generated according to the Revised UDP. This Appendix is intended to support statistical peer review of confidence interval procedures, and (subject to modifications based on the review) to support numerical programming.

Based on simulations presented in Section 2 of this document, it appears that in most cases it will be possible to compute a confidence interval using one of two procedures and that these procedures will have acceptable performance.

In cases where no animals respond at some doses, and all animals respond at some other doses (the latter being greater than the former), it is proposed that the lower bound for the LD50 will be the highest dose associated with no observed response. Similarly, the upper bound will be the lowest dose associated with responses for all animals tested at the dose.

In most other cases, it will be possible to compute a bound using the method of profile likelihood (see Barndorff-Nielsen, 1991, Section 10.2.4). In particular, it is proposed that this approach will be used in most cases where there is only one dose with an intermediate response fraction (neither 0% nor 100% responding), a case that is not handled by standard probit methods. (Proposals for handling various cases are summarized in Section 1.5 of this Appendix.)

Of the two procedures, the profile likelihood approach is the primary focus of this Appendix. The approach requires handling of a number of special cases and specification of other technical details.

Although a description of the profile likelihood approach has been included here, this document is intended to be reviewed primarily by individuals with some background in likelihood based statistical

procedures. In addition, it is assumed that readers are familiar with certain types of numerical techniques (line searching and optimization) as used in implementation of nonlinear statistical models.

The material which follows is organized into three sections.

Section 1.3 presents notation, the probit dose-response model, and the profile likelihood approach for computation of confidence intervals. Comments are provided on alternative parameterizations of the probit model.

Section 1.4 discusses numerical algorithms. Three types of specialized numerical routines are required: 2-dimensional optimization to calculate maximum-likelihood estimates, line searching to compute bounds, and 1-dimensional optimization (nested within the line search).

Section 1.5 presents a classification of cases, with proposals regarding how each case is to be handled. Different cases require different confidence interval computations and, for some low-frequency cases, confidence intervals are not proposed.

1.3 Overview of parametric approach

Notation for describing grouped data. For present purposes, it is convenient for the data to be summarized by dose level. Let:

g = number of dose levels tested;
d_i = ith dose level evaluated, $I = 1,...,g$. We assume that d_1 is lowest test dose, d_g is the highest, and so on.
x_i = $\log_{10}(d_i)$
n_i = number of animals tested at the ith dose level, $I = 1,...,g$;
r_i = number of animals observed to respond at the ith dose level, $I = 1,...,g$.

While data summarized in this way are convenient for the computations described here, some computations associated with the stopping rules cannot be calculated from data summarized in this way.

Probit dose-response model. A probit curve is fitted to the data, relating the fraction of animals that respond and the logarithm of dose. The probit model has two parameters. According to one parameterization (the parameterization proposed for final results), the probity parameters are the *slope* (say $) and the LD50. For purposes of this document, it is convenient to make use of the parameter μ=$\log_{10}$(LD50). For likelihood-based statistical procedures such as those used here, it is permissible to do estimates and confidence intervals directly for μ and then transform those results to results for the LD50.

Let $p(x ;\mu, \$)$ denote the probability of response, where x is the common logarithm of dose. Then an

expression for the probity model is:

$$p(x;,\$) = M[(x - \mu)\cdot\$]$$

where $M(z)$ denotes the cumulative distribution function (CDF) for a standard normal distribution.

Calling the parameter $\$$ a "slope" is a toxicological convention. Probity analysis is commonly described as a linear regression of a transformed response (probity percentage response) against the logarithm of dose. To see this, rearrange the expression above as follows:

$$M^{-1}[p(x;\mathrm{LD50},\$)] = \$\cdot x - \$\cdot\log_{10}(\mathrm{LD50}))$$

where M^{-1} denotes the inverse of function corresponding to M, so that evidently the relationship between dose and response can be transformed to a linear relationship with slope $\$$ and intercept -$\$\log_{10}(\mathrm{LD50})$).

Note the use here of the common (base-10) logarithm of dose, which is a toxicological convention. For some purposes, the choice of a base for logarithms is arbitrary, but the common logarithm needs to be used in software designed to support Revised UDP, in order to have comparability of results obtained with different programs. In particular, the value of the slope estimate will depend on the base chosen for logarithms.

An alternative parameterization, associated with a particular interpretation of the probity model, is:

$$p(x;\mu,F) = M[(x - \mu)/F]$$

where $\mu = \log_{10}(\mathrm{LD50})$ and $F = 1/\$$. Of course, μ and F^2 are conventional notation for the mean and variance of a normal distribution. This parameterization may be preferred particularly when the probity model is interpreted in terms of a *tolerance distribution*. According to that interpretation, variation among test animals in response to a particular dose is related to individual variation in sensitivity to the test substance. The tolerance of a single individual is defined to be the dose that will cause that individual to respond, given its sensitivity to the test substance. Then the fraction responding at a given dose equals the fraction of individuals with tolerance below that dose. A frequency distribution is assumed for variation of tolerances among individuals. The probity formulae result from assuming a lognormal distribution for tolerances, with parameters μ and F.

For purposes of the procedures described in this Appendix, the μ, $\$$ parameterization has proved to be more convenient than the μ, F parameterization. In particular, it appears that widely different values of F can be associated with slope values about equal to zero, and log-likelihood values that are not much different.

Point estimation of the LD50. This Appendix is concerned primarily with interval estimates rather

than with point estimates. However, the following remarks may help to place in perspective the various computations that need to be implemented in the software. The purpose of acute testing under the Revised UDP is to obtain an LD50 estimate. In this context, the probity slope is a nuisance parameter. Revised UDP specifies that when estimating the LD50, a value will be assumed for the slope parameter (the default assumption is a slope of 2) and that the LD50 will be estimated based on the resulting 1-parameter model using maximum likelihood. Revised UDP provides an expression for the likelihood function. The LD50 point estimate is not used in the computations for the confidence interval developed in this Appendix. Computations for the Revised UDP stopping rule also involve a distinct point estimate of the LD50, for different reasons.

Two-parameter and profile log-likelihood functions for grouped data. Likelihood functions are functions of model parameters, which are used in statistical inferences about those parameters. Each distinct data set yields a distinct likelihood function. It can be helpful to think of a likelihood function as measuring the relative support that the data provide for alternative choices of parameter values, with higher loglikelihood values indicating relatively stronger support. For example, the maximum-likelihood estimates of the parameters μ and $\$$ are the parameter values that maximize the 2-parameter function. The exact roles of these functions in computation of confidence intervals are described in detail below.

The following two likelihood functions need to be defined for the methods proposed. The log-likelihood function for the two-parameter probity model is:

$$l(m,b) = \sum_{i=1}^{g} \{ r_i \bullet \ln(p(x_i; m, b)) + (n_i - r_i) \bullet \ln(1 - p(x_i; m, b)) \}$$

(Note the use here of the natural logarithm rather than the common logarithm, which contrasts with the transformation of doses.)

Here, statistical inferences will focus on μ, whereas $\$$ will be treated as a nuisance parameter. In this context it is useful to define a type of loglikelihood that is a function of μ only, with $\$$ eliminated. The *profile loglikelihood* function is:

$$l_P(m) = \sup_b l(m,b)$$

In words, define the profile loglikelihood function to be the function of μ only, obtained by setting $\$$ equal to that value which maximizes the 2-parameter likelihood $l(\mu,\$)$, fixing μ. This requires a numerical optimization (numerical techniques are described in the next section). In practice the profile likelihood is handled using the same procedures as the likelihood of a single-parameter model, *e.g.*, in likelihood ratio tests (Barndorff-Nielsen, 1991).

Confidence intervals based on profile log-likelihood, "basic" approach. For the likelihood-based intervals considered here, the interval is the set of parameter values not rejected using a likelihood ratio test. The procedure can be stated most simply in the case where unique, finite maximum likelihood estimates (MLEs) exist for both probity parameters, in the interior of the space of allowable values. In this case the approach is fairly straightforward.

Let ^ be the MLE for μ and let $\hat{b}$ denote the ML for $, which is to say that $\hat{m}$ and $\hat{b}$ are the choices of parameter values that maximize the likelihood function. Then the *maximized value of the log-likelihood* , say $l_{\sup}$, is obtained by plugging the MLEs into the likelihood expression. Thus:

$$l_{\sup} = l(\hat{m}, \hat{b}) = l_P(\hat{m})$$

(Here "sup" is short for "supremum.") Then, for a 2-sided 95% confidence interval, the upper bound and lower bounds for μ, say $\overline{m}$ and $\underline{m}$, are obtained by solution of the following nonlinear equations:

$$l_P(\underline{m}) = l_P(\overline{m}) = l_{\sup} - 1.921, \quad \underline{m} < \hat{m} < \overline{m} .$$

In general, to compute a 100(1 - ")% confidence interval, the bounds are defined by the equation:

$$l_P(\underline{m}) = l_P(\overline{m}) = l_{\sup} - \frac{1}{2} c_1^2(1-a), \quad \underline{m} < \hat{m} < \overline{m}$$

(Bickel and Doksum, 2001) where $c_1^2(1-a)$ denotes the (1- ")th quantile of a chi-square distribution with a single degree of freedom. (In particular $c_1^2(0.95) = 3.84$=2*1.92.) It is useful to define

$$l_{\text{crit}} = l_{\sup} - \frac{1}{2} c_1^2(1-a)$$

which is the critical value of the profile loglikelihood that the bound values must satisfy.

Use of these expressions requires numerical searches among values of μ above and below $\hat{m}$. In some cases a solution does not exist, in which case the bound may be taken to be ± 4. In particular cases, graphs of the profile likelihood suggest an approach to an asymptote that falls short of the critical value. Unless conditions can be derived and automated for identifying the apparent infinite-bound cases, the search must be restricted to a finite interval. When the search is restricted to a finite interval, one cannot distinguish between bounds that are very wide and bounds that are actually infinite.

Example. The following hypothetical data were generated in a simulation of the Revised UDP. The profile loglikelihood curve for these data is displayed in Figure A.1

Table A.1. Data for Profile Loglikelihood Example

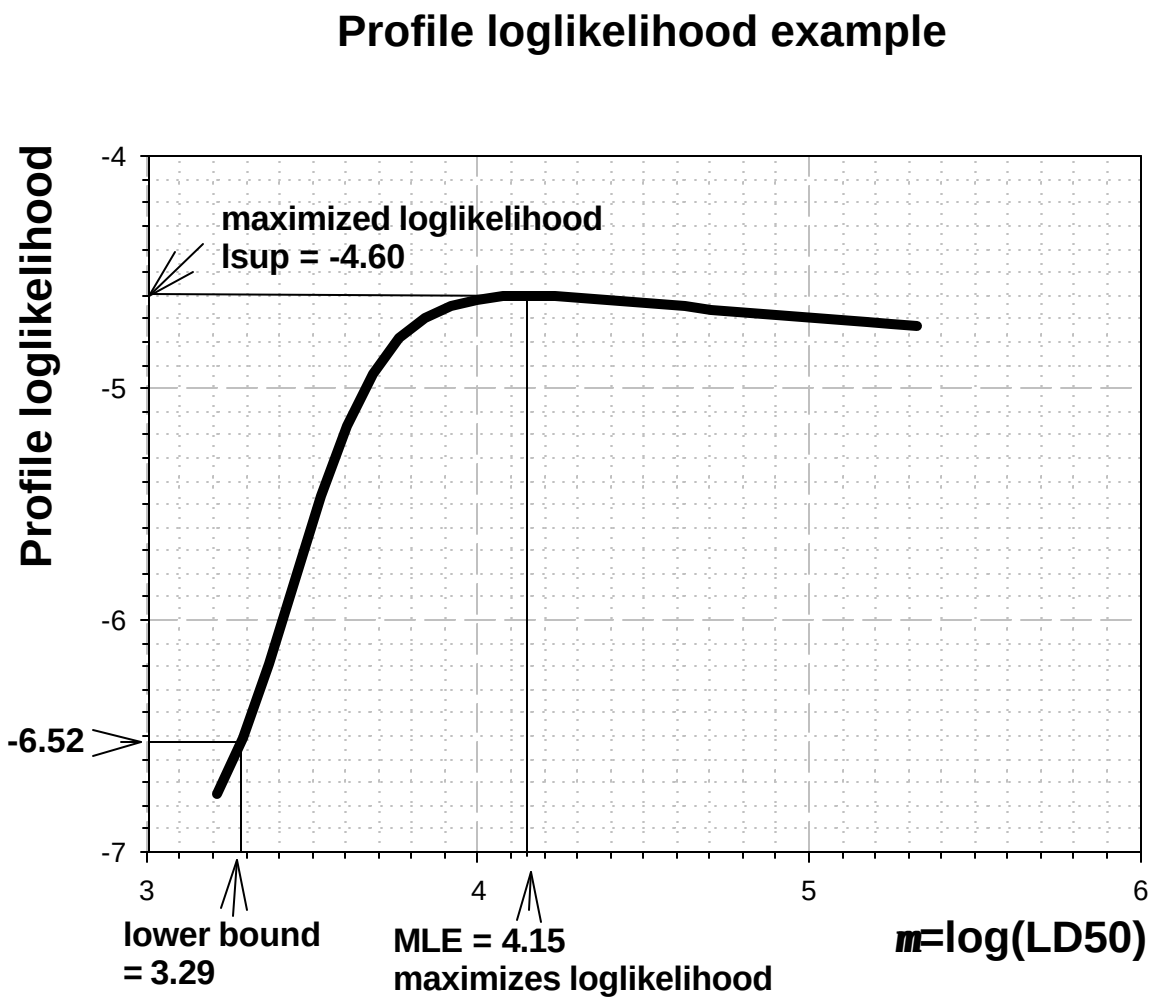
dose (mg/kg)	number tested	number responding
175.0	1	0
553.4	2	0
1750.0	3	1
5000.0	4	1

MLEs: $\hat{m} = 4.153$, $\hat{b} = 1.020$, estimated LD50=14223 mg/kg
95%CI for LD50 (1950 mg/kg,>2*10^5 mg/kg)
maximized loglikelihood: $l_{sup} = -4.603$
critical loglikelihood for bounds: l_{crit}=-6.524

This data set was analyzed in the following steps. The 2-parameter probit model was fitted to the data using a conventional probit methods (weighted Gauss-Newton optimization). That approach is considered to yield MLEs of model parameters in this case. The MLEs are displayed in table footnotes.

Evaluation of the loglikelihood at the MLEs gives l_{sup}=-4.603 (see Figure A.1). Therefore, any bounds must have profile loglikelihood equal to l_{crit}.=-6.524. A line search below the MLE found the lower bound for μ of 3.29 (or LD50=1950). A search for an upper bound failed to find a value of μ with the required profile loglikelihood within a factor of 15 of the MLE. Therefore, the upper bound would be reported as greater than 213000 (=15*14223). In this case, the absence of a useful upper bound probably results from the restriction of test doses to values not exceeding 5000 units

Figure A.1 Profile Loglikelihood Example



Extension of the approach to cases with a single intermediate response fraction. In some cases, the computations just described will not be applicable. However, the approach has been extended to one case that is not ordinarily analyzed under a 2-parameter probit model. This is the case where there is only a single test dose with an intermediate response fraction (the percentage responding is neither 0% nor 100%), and where any lower test doses are associated with 0% response, and any higher test doses are associated with 100% response. This is Case 4 as described in Table 1.

In the cases considered above, the loglikelihood supremum l_{sup} was found by evaluating the loglikelihood at the MLEs. For Case 4, it appears that l_{sup} has a natural definition, although the value of l_{sup} is obtained as a limit and does not correspond to particular finite values of μ and $\$$.

Within the ranges allowed for μ and $\$$, the fitted probit curve can be made to match the data as closely as we like by specifying $\$$ to be sufficiently steep. Consider the family of curves that exactly match the single partial response, with different slopes. Steeper slopes allow the 0's to be matched more closely at one end, and the 100's to be matched more closely at the other end. This argument suggests that the supremum of the loglikelihood can be calculated by taking appropriate limits, resulting in the expression:

$$l_{\text{sup}} = r_j \bullet \ln\left[\frac{r_j}{n_j}\right] + (n_j - r_j) \bullet \ln\left[\frac{n_j - r_j}{n_j}\right]$$

where j is the index of the dose associated with a partial response fraction. This expression is obtained from the 2-parameter log-likelihood $l(\mu,\$)$ by deleting the contributions from doses other than dose j, and for dose j by setting the predicting response percentage equal to the observed response percentage r_j / n_j. For the terms other than the jth, the limit is zero as the slope is taken to infinity. The jth observed response fraction equals the corresponding observed fraction because for any finite slope value, the intercept can be adjusted so that the results for the jth dose are matched exactly.

A second requirement for implementation of a profile likelihood approach is to define a finite interval of μ values in which to search for an upper or lower bound. Where there is an unambiguous MLE, an upper bound is searched for among values of μ above the MLE, and a lower bound is searched for below the MLE. In the case under consideration, where there is a single dose with partial response, we use the dose that has partial response as a bound for the search interval.

Example with one partial response dose. The following data were generated in a simulation of Revised UDP.

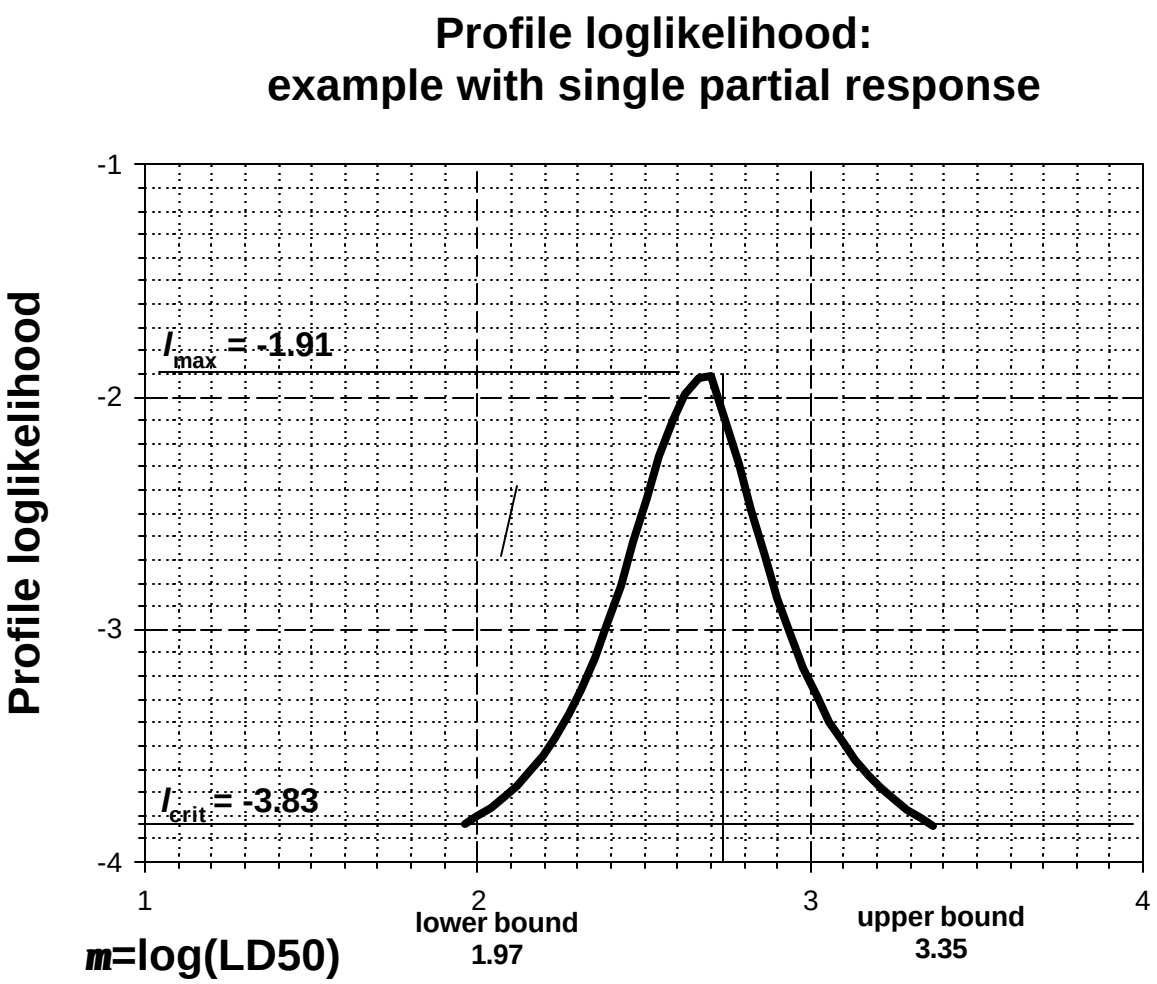
Table A.2. Example with a Single Partial Response Dose.

dose (mg/kg)	number tested	number responding
175.0	2	0
553.4	3	2
1750.0	1	1

maximized loglikelihood: l_{sup} = -1.910
critical loglikelihood for bounds: l_{crit}=-3.830
95%CI 1.97 - 3.35 for μ, 93 - 2258 mg/kg for the LD50

In this example, there is only a single partial response dose. The maximized loglikelihood is calculated using the formula given. (The LD50 would be 553.4.) The graph of the profile loglikelihood (Figure A.2) does not suggest any problem with this way of defining l_{sup}. Each point plotted corresponds to specific finite values of the parameters, but nevertheless the proposed method for calculating l_{sup} (which does not correspond to any particular parameter values) appears consistent with the rest of the curve. The use of such a profile loglikelihood presents no obvious problems.

Figure A.2. Profile Likelihood: Example with Single Partial Response.



1.4 Numerical Algorithms for Likelihood Calculations

A number of technical decisions are required in order to implement a profile likelihood procedure. It is desirable first of all to have criteria for determining if a 2-parameter maximization of the loglikelihood can be performed. In that case, the parameter values that maximize the loglikelihood are the MLEs. In any case, the computation of a bound for the LD50 requires a line search of a finite interval. Some procedure is needed to define the interval that will be searched for a bound. The line search involves evaluation of the profile loglikelihood function $l_P(\mu)$ for different values of μ. Each evaluation of the profile loglikelihood involves a
1-dimensional optimization ($ is optimized with μ fixed).

Each of these procedures requires a number of technical decisions. Most of these decisions are not related to the fundamental method, being more to production of a reliable algorithm. Here, a description is provided of the implementation used. In simulations, it appears that this algorithm never aborts because of numerical overflows or divisions by zero, etc. For concreteness, the procedure is described for computing the upper bound. The modifications needed for computation of a lower bound seem obvious in most cases.

Computation of MLEs by 2-dimensional optimization. When an optimum can be determined for the likelihood function, the results are used in calculating bounds by the profile likelihood method. There are many optimization techniques that can be considered for this purpose. In probit analysis, it is conventional to use a weighted Gauss-Newton approach devised by R.A. Fisher. This algorithm is described in Finney's (1971) Chapter 4. The approach is considered to generate maximum likelihood estimates in probit analysis. This algorithm is considered to be a perfectly good approach viewed from the standpoint of modern nonlinear statistical modeling. The algorithm is actually a special case of an approach widely used for generalized linear models, a broad class of nonlinear models (McCullagh and Nelder, 1989). The algorithm is closely related to the more familiar Newton-Raphson algorithm, but involves a simplified expression for the Hessian.

It is known that finite MLEs do not exist in some cases. Silvapulle (1981) has presented necessary and sufficient conditions for existence of MLEs for logit and probit models. The conditions are very general, addressing models with many regressors. In the case of probit analysis, the conditions apparently reduce to a requirement that some dose where one or more animals respond is lower than some other dose where one or more animals do not respond. A particular case of Silvalpulle's condition is the case where there are at least two doses with partial response fractions. The latter is sometimes used as a criterion for when probit analysis can be performed. Another case is when the observed relation between dose and fraction responding deviates from monotonicity.

Silvapulle's criterion allows an estimate if the probit slope equal to zero. If the slope is zero, the same response fraction is predicted at every dose. In that case either there is no estimate of the LD50 or else every dose is estimated to be the LD50. A great many applications of probit or logit models are not concerned with estimation of an LD50, and Silvapulle in particular does not discuss estimation of the

LD50.

Currently, 2-dimensional optimization is performed when the Silvapulle condition holds and when an additional criterion is met, which indicates a positive relationship between dose and observed response. (The handling of various cases is summarized in the following section.) In addition to the Silvapulle condition, a requirement is that the geometric mean dose is higher for animals that respond than for animals that do not respond. This condition is indicated in Revised UDP as a requirement for inferences regarding the μ.

Specification of interval searched for a bound. A numerical search for a bound for the LD50 must be restricted to a finite interval, particularly in view of the possibility that a bound may be infinite. The *search interval* is defined using two numbers, a point estimate and a multiplicative factor, say F_{search}. For computation of the upper bound, the search interval is

[LD50 point estimate, LD50 point estimate* F_{search}].

For the lower bound, the search interval is

[LD50 point estimate / F_{search}, LD50 point estimate].

With regard to notation, the usual practice of using μ (=$\log_{10}$(LD50)) instead of LD50 is deviated from. This is because, in the software, it is expected that all results will be expressed in terms of the LD50. The variable F_{search} will be accessible for modification for the user. Therefore, F_{search} is represented as a multiplicative factor applied to a point estimate of the LD50.

Here, use of the term "point estimate" is possibly the source of some confusion. The LD50 value which defines the search interval is not the LD50 point estimate indicated in Revised UDP. Therefore, the term "center of search interval" may be used in the remainder of the document. To avoid having to define additional symbols, $\hat{m}$ will continue to denote the center of the search interval although, in statistics, the "hat" (ˆ) over a parameter symbol ordinarily indicates a maximum likelihood estimate.

Determining if a bound exists within the search interval (bracketing step). The line search algorithm has two steps, a *bracketing step* and a *bisection step*. The bracketing step serves to determine whether a bound exists within the search interval. Also, the bracketing step produces quantities useful in the bisection step, which follows.

Expressing the model in terms of μ, the search interval for an upper bound can be denoted ($\hat{m}$, $\hat{m}$ + $\log_{10}$(F_{search})). A bound exists within the search interval provided that $l_P(\hat{m} + \log_{10}(F_{\text{search}})) < l_{\text{crit}}$. If this condition holds, then the bisection step can be used to locate the bound value within the search interval. Otherwise, the upper confidence bound is reported only as being greater than the bound of the search interval, i.e., as greater than

$\hat{m} + \log_{10}(F_{search})$.

This suggest that the bracketing step need only involve evaluation of the profile loglikelihood at the bounds of the search region. However, a more complex set of computations is used: Observe that evaluation of l_P involves optimization of $\$$. A starting estimate of $\$$ is required for each optimization. Therefore, it is reasonable to evaluate a sequence of μ values $\hat{m}$, $\hat{m}+*$, $\hat{m}+2*$, ..., where $*$ is some constant, stopping when the value of l_P is less than l_{crit} or the bound of the search region is attained. If this approach is used, then good starting values of $\$$ are usually available. The optimized value of $\$$ from one evaluation of l_P is a good starting value for use in the next optimization.

In simulations, $F_{search} = 50$ is used currently, and a value of $*$ is used such that the bound of the search region is attained in 40 steps.

Calculation of a bound by bisection. The use of bisection to calculate a bound for μ requires two values, say μ_1 and μ_2, that satisfy $l_P(\mu_1) > l_{sup}$ and $l_P(\mu_2) < l_{sup}$. Such values are provided by the final two values of μ evaluated in the bracketing step.

Gauss-Newton algorithm to optimize $\$$ with μ fixed. The profile loglikelihood function $l_P(\mu)$ is a function of μ obtained from the 2-parameter loglikelihood $l(\mu, \$)$ by optimizing $\$$, with μ fixed.

The Gauss-Newton approach, conventional for 2-dimensional optimization in probit analysis, is easily developed for the case of 1-dimensional optimization of $\$$. First, for the benefit of individuals familiar with generalized linear models, the probit model can be written in the following form:

$$M^{-1}[\,p(x;\mu,\$)\,] = (x - \mu)\,\$ = x^{*}\,\$.$$

where, as previously, x denotes the common logarithm of dose. From this it is evident that the 1-parameter model with μ fixed can be treated as a generalized linear model with a single regressor x^* (=$\log_{10}$(dose)-μ), with no intercept term, and with link function M^{-1} (the probit link). As usual in probit analysis, binomial variation is assumed at a given dose, which results in a factor of $p(1-p)$ in the regression weights.

The standard approach leads to the following scheme for updating the estimate of $\$$:

$$[\ \$ \text{ at (I+1)th iteration}\ ] = [\ \$ \text{ at ith iteration}\ ] + \delta_{\$}$$

where $\delta_{\$}$ can be computed with the expression:

$$d_b = \left(\frac{\partial}{\partial b} l(m,b) \right) \Big/ \left(\sum_{i=1}^{g} w_i x_i^{*2} \right)$$

with the quantities w_i, x_i, y_i defined in the following steps (recall definitions given already for d_i , r_i, and n_i):

x_i^*	$= \log_{10}(d_i) - \mu$	= value of "regressor" for ith treatment level, I=1,...,g;
Probit_i	$= x_i^* \cdot \beta$	= predicted probit value for ith treatment level;
p_i	$= M(\text{Probit}_i)$	= predicted response fraction at ith treatment level;
binV_i	$= p_i(1 - p_i) / n_i$	= binomial variance.
f_i	$= \exp(-\text{Probit}_i^2 / 2) / \sqrt{(2\pi)}$	= weight contribution associated with probit dose-response function;
w_i	$= f_i^2 / \text{binV}_i$	= weight for ith treatment level;
p_i^{obs}	$= r_i / n_i$	= observed response fraction at ith treatment level;

and

$$\frac{\partial}{\partial b} l(m,b) = \sum_{i=1}^{g} f_i \bullet x_i^* \bullet (p_i^{obs} - p_i) \bullet \text{binV}_i^{-1}$$

the last quantity being the partial of the 2-parameter loglikelihood with respect to β.

β is not constrained to be non-negative in these computations. An argument can be made for constraining β to be non-negative, or greater than some small positive value such as 0.5. Adding a constraint of this sort does not appear to be technically difficult, and would probably narrow some of the confidence intervals.

Convergence criteria. All that is needed from the 1-dimensional optimization is a profile-loglikelihood value. A relative gradient criterion can be used. Convergence occurs when

$$\frac{\partial}{\partial b} l(m,b) \Big/ l(m,b) \quad \# \; 0.00001.$$

For 2-dimensional optimization, a criterion based on relative change in parameter values is used currently.

Stabilization of parameter changes. When the starting values are too far from the optimum, the search direction indicated by the algorithm may be reasonable, while the magnitude of change in that direction may be such as to miss the optimum significantly. Improvements on the basic algorithm may

involve use of the search direction, with modification of the magnitude of change in that direction, for example by use of halving or line searching (Myers, 1990, particularly Section 9.4; Seber and Wild, 1989).

For the 1-dimensional optimizations, the magnitude of parameter change ($d_{\$}$) is constrained to absolute values not exceeding 0.5. $d_{\$}$ is set to 0.5 whenever $d_{\$}$ is greater than 0.5 and $d_{\$}$ is set to -0.5 whenever $d_{\$}$ is less than -0.5. This feature eliminated some problems that occurred otherwise.

Computation of starting values for optimizations. Convergence is expected to be rapid and reliable within a sufficiently small neighborhood of the optimum. Many authors emphasize computation of starting values likely to be close to the optimum solution. In the case of probit analysis, an obvious approach for computing starting values is by a linear regression of transformed response fraction (probit transformation) against log dose. The probit transformation is not finite valued if the response fraction is 0 or 1, hence a small constant may be added or subtracted from the observed response fractions, to obtain finite probit values for use in the regression.

A starting slope value is not calculated from the data when fitting the probit function. Experience with the standard Gauss-Newton algorithm used in probit analysis has shown that numerical failures may be associated with computation of weights. Note that the weight computations involve division by the quantities $p_i(1 - p_i)$ where p_i is the predicted response fraction at the ith treatment level based on the current parameter values. Numerical failures are often related to values of one or more of the p_i that are too close to 0 or 1, so that division by zero occurs. This outcome can be prevented by setting the initial value of the slope at a small value
(a value of 0.5 is used). For a starting value for the LD50, the geometric average of test doses is used.

A starting slope value from the data is not calculated when fitting the probit function. Instead, for a starting value for the LD50, the geometric average of test doses is used. Starting values of the slope are also needed for the 1-dimensional optimizations of $\$$ (fixing μ). For most of these optimizations, an optimized value of $\$$ corresponding to a nearby value of μ is available. Otherwise, a value of 0.5 can be used.

1.5 Classification of Cases

It is proposed that whether a confidence interval can be calculated, and if so the computations to be used, will be based on the following classification (see Table A.3).

In development of this scheme, discussions with the OECD avian stat group have been very helpful, although that group has developed a somewhat different classification (report in press). For example, the avian scheme does not explicitly use the results of Silvapulle.

The conditions for cases are checked in the order that the cases are displayed in the table, so when the conditions for a given case are met, none of the higher-number cases obtain. Table A.4

indicates the computational procedures proposed for each case. Subsequent text expands upon the suggestions summarized in this table.

Table A.3. Classification of Data Cases for Purposes of Confidence Interval Calculation

Case	Description
1	("No positive dose-response association"). There is no variation in response (all animals tested in the study responded, or none responded), or the geometric mean dose is lower for animals responding than for animals not responding.
2	("Standard 2-parameter probit computations"). The Silvapulle criterion holds (i.e., one or animals responded at a dose below some other dose where one or more did not respond. The conditions defining Case 1 do not hold.
3	("No partial response fractions."). All doses tested are associated with response fractions of 0% or 100%, with the doses associated with 0% response lower than the doses associated with 100% response. One or more doses is associated with 0% response and one or more doses is associated with 100% response.
4	("One partial response fraction, first subcase"). There is a single dose associated with a partial response fraction. One or more lower test doses is associated with 0% response, and one or more higher test doses is associated with 100% response.
5	("One partial response fraction, second subcase"). There is a single dose associated with partial response, which is either the highest test dose (with no responses at all other test doses) or the lowest test dose (with 100% response at all other test doses).

Acknowledgement. The development of this scheme was helped by conversations with the OECD avian acute statistics group, which has developed a similar classification (report in press).

Table A.4. Classification of Data Cases for Purposes of Confidence Interval Calculation with Computational Procedures

Case	Description	Confidence interval approach	2-parameter MLE calculated	Profile likelihood procedures	
				log-likelihood supremum	center of search region
1	No positive dose-response association	no confidence interval computed	no	not applicable	
2	Standard 2-parameter probit computations	basic profile likelihood approach	yes	equal to loglikelihood evaluated at the MLEs	MLE for LD50
3	No partial response fractions and not Case 1.	lower bound is highest with 0% response, etc.	no	not applicable	
4	One partial response fraction, 0% response at some lower doses and 100% at some higher doses	profile loglikelihood extended by special computations	no	expression in footnote[1]	Dose associated with partial response
5	One partial response fraction, at either high test dose or low test dose	profile loglikelihood extended by special computations	no	expression in footnote[1]	Dose associated with partial response

[1] Suppose the jth dose is associated with a partial response. Then the loglikelihood supremum is

$$l_{\sup} = r_j \bullet \ln\left[\frac{r_j}{n_j}\right] + (n_j - r_j) \bullet \ln\left[\frac{n_j - r_j}{n_j}\right]$$

where n_j and r_j denote the number of animals treated and the number that respond at the jth treatment level (see Section 1).

The decisions indicated in the table are as follows:

Case 1. With the stopping rules indicated for Revised UDP, this case appears to be possible only if testing is stopped at a limit dose (based on non-response for three animals tested in sequence at the dose). No methods are proposed here for cases where there is not an observable relationship between dose and response. In some cases, a binomial test may be used to establish that the LD50 is above or below the range of doses tested, but a significant binomial test requires testing of 5 or more animals and would use only the data from one test dose. Some procedures that may be applicable in this case have been developed for avian acute testing (report in press).

Case 2. Where the data allow, both probit parameters are estimated using maximum likelihood. The loglikelihood supremum is the value of the 2-parameter loglikelihood, evaluated at the MLEs.

This loglikelihood supremum is used to calculate a critical loglikelihood, which the bound values must satisfy. A search above the LD50 MLE is used to calculate an upper bound and a search below the LD50 MLE is used to calculate a lower bound.

Case 3. When there are no partial response fractions (along with other requirements of the case, as indicated in Table A.4) the profile loglikelihood approach apparently cannot be used. In this case, it seems that any dose within the interval bounded by the highest dose with no response, and the lowest dose with 100% response, would be equally valid as an estimate of the LD50. Simulations suggest that these two doses will perform acceptably when used as confidence bounds. Graphs of the profile loglikelihood indicate discontinuities at those doses, so that the profile loglikelihood approach cannot be implemented in a straightforward manner.

Case 4. When there is a single partial response (along with other requirements for the case, as indicated), the profile loglikelihood can be applied using special computations as described in Section 1. It is proposed that, when searching numerically for a bound, the dose with partial response can be used to define the search interval.

Case 5. This is an infrequent case which occurs mainly if the LD50 is close to a bound.

1.6 References

Barndorff-Nielsen, O.E. 1991. Likelihood theory. Chapter 10 in D.V. Hinkley, N. Reid, and E.J. Snell (eds) *Statistical Theory and Modelling*. Chapman and Hall.

Bickel, E.J. and K. A. Doksum. 2001. *Mathematical Statistics: basic ideas and selected topics*. Volume 1. (2nd ed.) Prentice Hall.

Finney, D.J. 1971. *Probit Analysis*. 3rd ed. Cambridge U. Press.

McCullagh, P., and J.A. Nelder. 1989. *Generalized linear models*. (2nd ed.) Chapman & Hall/CRC.

Myers, R.H. 1990. *Classical and modern regression with applications*. (2nd ed.) Duxbury Press.

Seber, G.A.F., and Wild, C.J. 1989. *Nonlinear regression*. John Wiley and Sons.

Silvapulle, M.J. 1981. On the existence of a maximum likelihood estimators for the binomial response model. J. Royal Statist. Soc. Series B 43(3):310-313.

Appendix B

Tables of Simulation Results

Table B.1. Percentages of Cases 1-5 among Simulated Studies

Scenario#	LD50	slope	initial test dose	% Case 1 + Case 5	% Case 2	% Case 3	% Case 4
(I) Scenarios with initial test dose 175 units							
1	1.5	8.33	175	21.3	0.0	78.7	0.0
2		4	175	53.1	3.0	42.6	1.3
3		2	175	41.6	31.0	18.5	8.9
4		0.8	175	19.7	61.6	6.0	12.7
5		0.5	175	11.5	67.7	5.7	15.1
6	2.5	8.33	175	0.0	0.0	99.2	0.8
7		4	175	10.8	6.3	64.7	18.2
8		2	175	13.9	38.5	21.9	25.8
9		0.8	175	10.6	66.3	6.7	16.5
10		0.5	175	9.3	70.4	5.1	15.2
11	20	8.33	175	0.0	0.0	35.3	64.7
12		4	175	0.0	9.3	24.9	65.8
13		2	175	0.0	40.6	14.8	44.6
14		0.8	175	0.2	61.9	7.6	30.3
15		0.5	175	1.5	61.7	7.2	29.7
16	50	8.33	175	0.0	0.0	29.8	70.2
17		4	175	0.0	7.3	24.0	68.7
18		2	175	0.0	37.2	12.4	50.4
19		0.8	175	0.0	54.9	8.4	36.7
20		0.5	175	0.3	57.5	7.0	35.2
21	150	8.33	175	0.0	0.0	36.7	63.3
22		4	175	0.0	4.1	26.6	69.3
23		2	175	0.0	26.1	15.8	58.1
24		0.8	175	0.0	48.5	9.8	41.7
25		0.5	175	0.6	56.6	8.0	34.9
26	600	8.33	175	0.0	0.0	30.3	69.7
27		4	175	0.0	6.7	22.9	70.4
28		2	175	0.0	32.6	12.7	54.7
29		0.8	175	0.6	54.3	9.0	36.1
30		0.5	175	1.9	58.9	8.6	30.6
31	1500	8.33	175	0.0	0.0	39.9	60.1
32		4	175	0.2	9.3	24.8	65.8
33		2	175	1.2	43.4	13.5	41.9
34		0.8	175	4.4	59.8	6.5	29.3
35		0.5	175	6.0	62.1	5.8	26.0
36	3000	8.33	175	9.5	1.1	82.4	7.0

Scenario#	LD50	slope	initial test dose	% Case 1 + Case 5	% Case 2	% Case 3	% Case 4
37		4	175	21.0	25.4	30.0	23.5
38		2	175	14.7	52.4	11.3	21.6
39		0.8	175	11.2	62.9	6.4	19.5
40		0.5	175	11.2	60.1	5.2	23.5
41	3500	8.33	175	27.4	1.0	70.5	1.1
42		4	175	36.1	24.9	28.0	11.1
43		2	175	22.4	50.9	9.5	17.2
44		0.8	175	12.1	62.6	6.4	18.9
45		0.5	175	11.0	60.0	6.8	22.3
(ii) Scenarios with initial test dose varied							
46	1.5	8.33	1.1	0.0	2.1	66.0	31.9
47			1.5	1.2	9.2	22.0	67.6
48			1.9	8.0	9.3	43.1	39.6
49		4	1.5	4.2	27.6	16.1	52.1
50			2.4	18.6	27.8	23.9	29.7
51		2	1.5	9.1	40.6	12.3	38.0
52			4	30.9	39.0	14.1	16.0
53		0.8	1.5	15.5	52.5	6.2	25.9
54			16.9	19.5	58.7	6.5	15.2
55		0.5	1.5	19.3	50.0	6.7	24.0
56			72.3	8.2	67.4	5.8	18.6
57	2.5	8.33	1.8	0.0	0.1	67.6	32.3
58			2.5	0.0	0.0	26.1	73.9
59			3.1	0.0	0.0	50.1	49.9
60		4	1.2	0.0	10.1	33.4	56.5
61			2.5	0.7	8.2	22.6	68.4
62			4.1	6.5	8.9	43.3	41.3
63		2	2.5	3.1	38.4	14.1	44.5
64			6.6	1.5	40.0	13.6	44.9
65		0.8	2.5	11.8	53.3	7.0	28.0
66			28.2	6.8	60.4	7.5	25.3
67		0.5	2.5	14.0	54.3	6.4	25.4
68			120.5	7.1	67.3	5.9	19.6
69	20	8.33	14	0.0	0.2	74.1	25.7
70			20	0.0	0.0	25.7	74.3
71			25.2	0.0	0.0	50.0	50.0
72		4	9.6	0.0	9.5	34.0	56.5
73			20	0.0	5.0	21.7	73.3
74			32.5	0.0	10.8	34.3	54.9

Scenario#	LD50	slope	initial test dose	% Case 1 + Case 5	% Case 2	% Case 3	% Case 4
75		2	4.6	0.0	41.1	14.6	44.3
76			20	0.0	28.1	16.6	55.3
77			52.7	0.0	32.3	14.3	53.4
78		0.8	20	0.1	51.2	8.9	39.8
79			225.4	0.1	62.4	7.8	29.7
80		0.5	20	0.9	59.5	8.2	31.4
81			964.4	1.5	71.6	6.3	20.6
82	50	8.33	35.1	0.0	0.0	73.9	26.1
83			50	0.0	0.0	22.6	77.4
84			63.1	0.0	0.0	50.8	49.2
85		4	23.9	0.0	9.2	36.1	54.7
86			50	0.0	3.3	20.9	75.8
87			81.2	0.0	8.8	34.8	56.4
88		2	11.4	0.0	35.6	15.0	49.4
89			50	0.0	27.4	14.2	58.4
90			131.8	0.0	32.1	13.5	54.4
91		0.8	1.3	0.0	68.1	7.6	24.3
92			50	0.0	51.8	8.0	40.2
93			563.6	0.0	58.7	8.6	32.7
94		0.5	50	0.8	57.4	8.2	33.5
95			2411.1	1.5	69.6	7.2	21.8
96	150	8.33	105.3	0.0	0.0	71.8	28.2
97			150	0.0	0.0	24.6	75.4
98			189.3	0.0	0.0	50.0	50.0
99		4	71.7	0.0	9.5	33.0	57.5
100			150	0.0	4.9	21.4	73.7
101			243.5	0.0	9.3	34.7	56.0
102		2	34.3	0.0	36.0	14.5	49.5
103			150	0.0	26.7	16.8	56.5
104			395.3	0.0	32.0	13.6	54.4
105		0.8	3.8	0.2	70.0	5.4	24.5
106			150	0.0	51.5	9.1	39.4
107			1690.9	0.3	62.0	8.0	29.7
108		0.5	150	0.7	55.7	8.2	35.4
109	600	8.33	421	0.0	0.1	72.7	27.2
110			600	0.0	0.0	26.9	73.1
111			757.2	0.0	0.1	51.4	48.5
112		4	286.9	0.0	11.4	33.7	54.9
113			600	0.0	4.3	25.3	70.4

Scenario#	LD50	slope	initial test dose	% Case 1 + Case 5	% Case 2	% Case 3	% Case 4
114			974	0.0	13.2	35.8	51.0
115		2	137.2	0.0	36.8	14.8	48.4
116			600	0.1	26.7	16.3	56.9
117			1581.1	0.4	31.8	14.1	53.7
118		0.8	15	0.1	69.6	7.1	23.2
119			600	1.2	52.9	8.4	37.5
120		0.5	1.6	1.5	75.5	5.1	17.9
121			600	3.0	59.6	6.4	31.0
122	1500	8.33	1052.5	0.0	0.0	72.9	27.1
123			1500	0.0	0.0	23.4	76.6
124			1892.9	0.0	0.0	52.4	47.6
125		4	717.3	0.0	7.5	34.2	58.3
126			1500	0.2	5.1	23.6	71.2
127			2435	0.0	9.4	34.6	56.0
128		2	343	3.8	39.8	17.2	39.2
129			1500	3.0	30.3	14.2	52.6
130			3952.8	0.5	37.2	13.9	48.4
131		0.8	37.5	4.7	69.0	6.2	20.1
132			1500	11.0	51.8	7.4	29.8
133		0.5	4.1	5.2	74.6	5.2	15.0
134			1500	15.1	52.4	6.6	25.9
135	3000	8.33	2105.1	5.4	2.9	66.4	25.3
136			3000	0.8	4.9	23.2	71.1
137			3785.8	0.2	1.5	52.7	45.6
138		4	1434.6	27.3	14.9	39.7	18.1
139			3000	3.0	24.8	20.1	52.1
140			4750	0.7	17.7	35.1	46.6
141		2	686	11.5	46.3	11.9	30.3
142			3000	8.5	40.5	11.8	39.2
143		0.8	75	7.9	67.5	5.4	19.2
144			3000	14.0	52.7	4.8	28.5
145		0.5	8.2	5.5	76.4	4.5	13.6
146			3000	18.2	52.9	6.9	22.0
147	3500	8.33	2455.9	17.8	6.6	53.3	22.3
148			3500	1.9	13.4	19.8	64.9
149			4416.8	0.0	4.8	50.5	44.7
150		4	1673.7	37.8	18.0	28.3	15.9
151			3500	4.7	30.4	16.0	48.9
152		2	800.4	13.9	50.0	9.0	27.1

Scenario#	LD50	slope	initial test dose	% Case 1 + Case 5	% Case 2	% Case 3	% Case 4
153			3500	8.1	43.4	11.2	37.3
154		0.8	87.5	9.2	66.8	6.0	18.0
155			3500	15.5	52.4	5.8	26.2
156		0.5	9.6	13.3	69.3	5.1	12.3
157			3500	18.1	54.2	5.2	22.4

Table B.2. Performance Measures PM1-PM4 (defined in Section 2.5).

Scenario #	LD50	slope	initial test dose	PM1(%)	PM2(%)	PM3	PM4(%)
(I) Scenarios with initial test dose of 175 units							
1	1.5	8.33	175	100.0	100.0	5.5	302
2		4	175	98.4	99.6	5.5	302
3		2	175	76.4	93.8	10.5	449
4		0.8	175	53.8	87.2	>1000	3033
5		0.5	175	45.2	79.6	>1000	>100000
6	2.5	8.33	175	100.0	100.0	5.5	181
7		4	175	99.6	99.9	5.5	181
8		2	175	89.7	96.5	7.0	275
9		0.8	175	58.2	88.1	>1000	2167
10		0.5	175	46.6	80.9	>1000	>100000
11	20	8.33	175	100.0	96.0	4.2	178
12		4	175	99.1	92.7	4.2	178
13		2	175	88.2	89.0	8.8	213
14		0.8	175	58.0	77.4	260.5	3425
15		0.5	175	52.5	73.2	>1000	5029
16	50	8.33	175	100.0	95.3	4.0	118
17		4	175	97.0	90.7	4.4	185
18		2	175	75.2	88.8	11.1	269
19		0.8	175	56.8	85.6	89.4	2012
20		0.5	175	52.2	81.8	>1000	4332
21	150	8.33	175	100.0	97.8	24.5	457
22		4	175	95.9	93.9	24.5	457
23		2	175	74.1	88.7	24.5	457
24		0.8	175	56.3	80.6	24.5	1250
25		0.5	175	50.0	79.1	>1000	>100000
26	600	8.33	175	100.0	93.8	4.0	191
27		4	175	96.9	89.2	4.2	191
28		2	175	77.8	89.0	10.5	224
29		0.8	175	55.3	84.0	63.3	4092
30		0.5	175	48.2	81.0	>1000	>100000
31	1500	8.33	175	100.0	97.1	4.1	135
32		4	175	98.8	93.0	4.1	214
33		2	175	82.6	89.0	10.3	247
34		0.8	175	51.7	79.8	>1000	>100000
35		0.5	175	44.7	76.9	>1000	>100000
36	3000	8.33	175	99.8	100.0	2.9	108

Scenario #	LD50	slope	initial test dose	PM1(%)	PM2(%)	PM3	PM4(%)
37		4	175	93.4	98.4	3.6	108
38		2	175	73.0	93.9	14.2	574
39		0.8	175	46.6	81.2	>1000	>100000
40		0.5	175	43.9	75.3	>1000	>100000
41	3500	8.33	175	99.7	100.0	2.9	93
42		4	175	90.2	99.5	3.6	93
43		2	175	64.0	94.6	108.9	1296
44		0.8	175	48.0	83.0	>1000	>100000
45		0.5	175	45.3	75.5	>1000	>100000

(ii) Scenarios with initial test dose varied

46	1.5	8.33	1.1	97.9	99.9	3.2	159
47			1.5	91.9	100.0	5.7	216
48			1.9	93.3	98.8	9.2	332
49		4	1.5	73.0	99.1	12.4	441
50			2.4	74.5	98.7	14.4	510
51		2	1.5	57.2	94.1	12.4	441
52			4	59.6	97.1	16.5	702
53		0.8	1.5	42.8	90.8	>1000	>100000
54			16.9	40.5	81.6	>1000	>100000
55		0.5	1.5	43.7	86.2	>1000	>100000
56			72.3	46.9	74.7	>1000	>100000
57	2.5	8.33	1.8	99.9	99.9	3.2	156
58			2.5	100.0	100.0	15.6	329
59			3.1	100.0	99.2	3.2	268
60		4	1.2	90.0	97.3	4.4	192
61			2.5	92.2	99.4	15.6	329
62			4.1	94.0	99.6	5.8	224
63		2	2.5	63.4	96.3	19.9	532
64			6.6	63.6	94.4	24.5	1033
65		0.8	2.5	44.4	88.3	>1000	>100000
66			28.2	52.6	79.5	>1000	4415
67		0.5	2.5	42.5	87.0	>1000	>100000
68			120.5	46.8	77.4	>1000	>100000
69	20	8.33	14	99.8	100.0	3.2	151
70			20	100.0	100.0	24.5	391
71			25.2	100.0	99.3	3.2	272
72		4	9.6	90.9	97.7	4.4	213
73			20	95.0	98.9	24.5	391

Scenario #	LD50	slope	initial test dose	PM1(%)	PM2(%)	PM3	PM4(%)
74			32.5	89.2	99.0	24.5	295
75		2	4.6	80.7	87.7	10.8	317
76			20	72.6	93.7	24.5	575
77			52.7	74.4	90.6	24.5	479
78		0.8	20	52.7	85.2	63.2	2288
79			225.4	63.6	76.6	70.2	1125
80		0.5	20	47.0	80.8	>1000	>100000
81			964.4	56.7	77.3	>1000	4874
82	50	8.33	35.1	100.0	99.9	3.2	152
83			50	100.0	100.0	24.5	391
84			63.1	100.0	99.1	3.2	86
85		4	23.9	91.5	96.8	4.4	183
86			50	96.7	99.1	24.5	391
87			81.2	91.2	98.7	24.5	295
88		2	11.4	81.5	89.1	10.9	282
89			50	72.7	91.0	24.5	575
90			131.8	74.0	90.7	24.5	479
91		0.8	1.3	72.8	77.0	81.4	2301
92			50	54.4	84.0	63.3	1238
93			563.6	66.9	73.9	33.2	973
94		0.5	50	48.7	79.3	>1000	>100000
95			2411.1	58.9	75.8	>1000	4121
96	150	8.33	105.3	100.0	99.9	3.2	152
97			150	100.0	100.0	24.5	391
98			189.3	100.0	99.7	3.2	273
99		4	71.7	90.7	97.1	4.4	251
100			150	95.1	98.9	24.5	391
101			243.5	90.8	98.9	24.5	295
102		2	34.3	83.0	91.4	8.8	272
103			150	73.5	92.0	24.5	575
104			395.3	72.2	90.7	24.5	479
105		0.8	3.8	70.0	78.6	120.2	3826
106			150	53.3	84.0	64.5	1238
107			1690.9	62.7	76.3	75.8	1139
108		0.5	150	50.5	80.0	>1000	25569
109	600	8.33	421	100.0	99.9	3.2	152
110			600	100.0	100.0	24.5	391
111			757.2	99.9	99.3	3.2	86
112		4	286.9	92.5	97.0	4.9	207

Scenario #	LD50	slope	initial test dose	PM1(%)	PM2(%)	PM3	PM4(%)
113			600	95.7	98.7	24.5	391
114			974	90.7	99.2	4.8	219
115		2	137.2	80.8	91.1	9.1	281
116			600	73.3	92.6	24.5	575
117			1581.1	74.9	91.7	24.5	266
118		0.8	15	65.1	81.2	183.2	3419
119			600	50.8	84.4	>1000	1509
120		0.5	1.6	62.2	78.7	>1000	>100000
121			600	45.5	82.5	>1000	>100000
122	1500	8.33	1052.5	100.0	100.0	4.8	263
123			1500	100.0	100.0	25.6	405
124			1892.9	100.0	99.3	3.2	86
125		4	717.3	92.6	98.1	4.0	166
126			1500	95.0	99.6	25.6	405
127			2435	90.6	98.9	10.7	295
128		2	343	74.6	94.0	6.7	261
129			1500	72.7	96.1	27.0	617
130			3952.8	69.2	87.7	10.5	358
131		0.8	37.5	59.3	84.0	>1000	63283
132			1500	46.0	90.8	>1000	>100000
133		0.5	4.1	51.7	78.5	>1000	>100000
134			1500	45.4	87.3	>1000	>100000
135	3000	8.33	2105.1	98.4	99.9	2.4	96
136			3000	95.6	100.0	7.1	225
137			3785.8	98.7	99.1	3.2	86
138		4	1434.6	89.3	99.9	3.5	119
139			3000	77.0	99.3	13.7	256
140			4750	82.8	99.1	4.4	137
141		2	686	58.6	93.2	24.5	231
142			3000	57.2	95.0	13.7	256
143		0.8	75	51.8	82.9	>1000	>100000
144			3000	42.2	90.1	>1000	>100000
145		0.5	8.2	52.3	79.9	>1000	>100000
146			3000	42.1	85.7	>1000	>100000
147	3500	8.33	2455.9	94.7	99.3	2.0	73
148			3500	87.4	100.0	5.2	172
149			4416.8	95.3	99.4	3.2	86
150		4	1673.7	83.4	99.9	9.8	254
151			3500	69.7	99.4	11.8	229

Scenario #	LD50	slope	initial test dose	PM1(%)	PM2(%)	PM3	PM4(%)
152		2	800.4	53.7	94.6	864.9	23232
153			3500	54.5	96.1	24.5	229
154		0.8	87.5	56.3	80.4	>1000	>100000
155			3500	40.8	89.8	>1000	>100000
156		0.5	9.6	46.6	81.1	>1000	>100000
157			3500	39.7	86.6	>1000	>100000

DESCRIPTION OF THE ACUTE ORAL TOXICITY SOFTWARE PROGRAM

Under contract to the U.S. EPA, Westat (Rockville, MD) developed the Acute Oral Toxicity (Guideline 425) Statistical Program" (AOT425StatPgm) to perform the statistical calculations associated with the OECD GUIDELINE FOR THE TESTING OF CHEMICALS, Section 4: Health Effects Test No. 425, Acute Oral Toxicity: Up-and-Down Procedure (OECD TG 425). The program may also be used with the U.S. EPA guideline, "Acute Oral Toxicity: Up-and-Down Procedure". The AOT425StatPgm program recommends a dose level for each animal, determines when dosing should stop, and calculates point and interval estimates for the LD50.

In addition to developing the AOT425StatPgm, Westat performed quality assurance testing for the statistical program and developed a report called, "QA Testing for the AOT425StatPgm Program." This document presents the results of testing to assess and to document the quality of the calculations performed by the program.

Westat also provided a complete report, called "Simulation Results for the AOT425StatPgm Program", which presents graphs of estimated LD50s, confidence interval (CI) widths, and expected animal usage, all obtained by automated means. These simulations were carried out to gauge the performance of the program and to determine the statistical performance of the OECD TG 425 procedure under various conditions.[1] They examine both the Main Test and the Limit Test.

The Acute Oral Toxicity 425 Statistical program (AOT425StatPgm), the quality assurance report, and simulations report which were made available to the UDP Peer Review Panel (Panel) for the August 21, 2001 review are currently found on the Internet at: http://iccvam.niehs.nih.gov/methods/udpdocs/udprpt/udp_ciprop.htm.

The software program and accompanying documentation are currently being revised by the U.S. EPA in accordance with recommendations from the Panel and comments received from the public. Upon completion of these revisions, the software program and accompanying documentation will be available through the U.S. EPA, Office of Pesticide Programs website at: http://www.epa.gov/pesticides/.

---

[1] The simulations assumed that the long-term and short-term outcomes (death or survival) were the same. One thousand simulated tests were generated for each simulation scenario, which included specified test type (Limit or Main), limit dose (2000 or 5000 mg/kg), and other defining features of the dose-response profile and its evaluation (e.g., true LD50 and sigma; assumed LD50 and sigma).

EVALUATION GUIDANCE TO THE PEER REVIEW PANEL FOR THE REVISED UP AND DOWN PROCEDURE (UDP)

Review of the Revised Draft UDP Test Guideline, a Procedure for Calculating Confidence Intervals for the LD50, and a Software Program for use with the UDP

Instructions for the Peer Review Panel

The Panel is asked to review: 1) the revised draft UDP; 2) the proposed procedure for calculating the confidence interval for the LD50; and 3) the UDP software program. In reviewing these documents/products, the Panel is asked to consider all of the information provided and develop written responses to the questions below. The Panel is asked to prepare a written report summarizing their review.

This review will focus on the following:

1. The revised draft UDP test guideline (June 20, 2001) incorporates modifications in accordance with the Panel's recommendations at the July 25, 2000 Peer Review Panel meeting.
   a) Are the changes consistent with the Panel's recommendations?
   b) Do you concur with the revisions that have been made?

   Note: This revised draft UDP test guideline (GUIDELINE FOR THE TESTING OF CHEMICALS: Acute Oral Toxicity: Revised Up-and-Down Procedure. Draft, July 12, 2001) was developed by UDP Technical Task Force and submitted to ICCVAM on July 12, 2001. To allow comparison with the version reviewed by the Panel (Revised Test Guideline 425N, April 11, 2000), a Summary of Changes document is being distributed to the Panel.

2. Is the proposed procedure for calculating a confidence interval for the LD50 appropriate and adequate for use with the revised draft UDP test guideline?

   Note: This document, "Description of Performance and Confidence Intervals for the Revised Up-and-Down Procedure (UDP) for Acute Oral Toxicity. June 6, 2001," was developed by the U.S. EPA and submitted to ICCVAM for distribution to the Panel members.

3. Is the software program adequate and consistent with the procedures in the revised draft UDP test guideline?

   Note: This software program and accompanying user's manual (Acute Oral Toxicity (Guideline 425) Statistical Program. May 2001) were developed by Westat, Inc. for the U.S. EPA and submitted to ICCVAM for distribution to the Panel members.

APPENDIX D

Federal Register Notices

D–1 Vol. 65, No. 34, February 18, 2000 .....D-3
Request for Data and Nomination of Expert Scientists

D–2 Vol. 65, No. 106, June 1, 2000 .....D-5
Notice of Peer Review Panel Meeting and Request for Comments

D–3 Vol. 66, No. 121, June 22, 2001 .....D-7
Notice of Availability and Request for Comments

D–4 Vol. 66, No. 133, July 21, 2001 .....D-11
Notice of Peer Review Panel Meeting

of Technology Transfer, National Institutes of Health, 6011 Executive Boulevard, Suite 325, Rockville, MD 20852–3804; Telephone: (301) 496–7056, ext. 211; Facsimile: (301) 402–0220.

SUPPLEMENTARY INFORMATION: Uteroglobin plays a significant role in human renal disease through its effect on the deposition of IgA. This invention relates to the use of uteroglobin and its role in the diagnosis and treatment of IgA nephropathy.

The prospective exclusive license will be royalty-bearing and will comply with the terms and conditions of 35 U.S.C. 209 and 37 CFR 404.7. The prospective exclusive license may be granted unless, within 90 days from the date of this published Notice, NIH received written evidence and argument that establishes that the grant of the license would not be consistent with the requirements of 35 U.S.C. 209 and 37 CFR 404.7.

The field of use may be limited to the use of the invention for the development of therapeutic and diagnostic applications relating to IgA nephropathy.

Properly filed competing applications for a license filed in response to this notice will be treated as objections to the contemplated license. Comments and objections submitted in response to this notice will not be made available for public inspection, and, to the extent permitted by law, will not be released under the Freedom of Information Act, 5 U.S.C. 552.

Dated: February 14, 2000.

Jack Spiegel,

Director, Division of Technology Development and Transfer, Office of Technology Transfer.

[FR Doc. 00–4009 Filed 2–17–00; 8:45 am]

BILLING CODE 4140–01–M

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service

National Institute of Environmental Health Sciences, National Toxicology Program: Request for Data and Nomination of Expert Scientists To Participate in the Independent Peer Review Evaluation of the Revised Up-and-Down Procedure for Assessing Acute Oral Toxicity; Evaluation of the Up-and-Down Procedure

The Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) are currently planning a meeting where an Independent Peer Review Panel (hereafter, Panel) will assess the validation status of the revised Up-and-Down Procedure (UDP). This procedure is an updated version of the Organization for Economic Cooperation and Development (OECD) Test Guideline 425 (OECD Guideline for the Testing of Chemicals, Acute Oral Toxicity: Up-and-Down Procedure. Guideline 425, adopted September 21, 1998, OECD, Paris, France, *http://www.oecd.org/ehs/test*). The revised UDP is proposed as a substitute for the existing OECD Test Guideline 401 (OECD Guideline for the Testing of Chemicals, Acute Oral Toxicity, Guideline 401, adopted February 24, 1987, OECD, Paris, France). OECD has proposed that Guideline 401 should be deleted since three alternative methods are not available (OECD Document ENV/JM(99)19, Test Guidelines Programme, Acute Oral Toxicity Testing: Data Needs and Animal Welfare Considerations, 29th Joint Meeting, June 8–11, 1999, Paris, France). Prior to deletion of Guideline 401, U.S. agencies have requested that ICCVAM conduct an independent peer review of the revised UDP to determine the validity of the method as a replacement for Guideline 401. The Panel will evaluate the extent to which the validation and acceptance criteria (outline in NIH Publication 97–3981, Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods, *http://ntpserver.niehs.nih.gov/htdocs/ICCVAM/iccvam.html*) have been addressed and will provide conclusions and recommendations regarding the usefulness and limitations of the method as a substitute for the traditional acute oral toxicity test method (OECD Guideline 401, 1987). The UDP has the potential to reduce the number of animals required to classify chemicals for acute oral toxicity as compared to Guideline 401.

Nomination of Experts To Serve on Review Panel and Request for Data

The Center welcomes the nomination of scientists with relevant knowledge and experience who might be considered for the Panel to review information on UDP. For each person suggested, his/her name, address, and a brief summary of relevant experience and qualifications should be provided. Where possible, telephone and fax numbers and/or e-mail address should also be provided. Nominations should be sent by mail, fax, or e-mail to NICEATM within 30 days of this notice's publication date. Correspondence should be directed to Dr. William S. Stokes, Co-Chair, ICCVAM, NTP Interagency Center for the Evaluation of Alternative Toxicological Methods, Environmental Toxicology Program, NIEHS/NTP, 79 T.W. Alexander Drive, MD EC–17, P.O. Box 12233, Research Triangle Park, NC 27709; phone: 919–541–7997; fax: 919–541–0947; e-mail: *iccvam@niehs.nih.gov.*

The Center would also welcome data and information from completed, ongoing, or planned studies using or evaluating the UDP. Information should address applicable aspects of the validation and regulatory acceptance criteria provided in NIH Publication 97–3981, Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods (*http://ntp-server.niehs.nih.gov/htdocs/ICCVAM/iccvam.html*). Where possible, data and information should adhere to the guidance provided in the document, Evaluation of the Validation Status of Toxicological Methods: General Guidelines for Submissions to ICCVAM (*http://iccvam.niehs.nih.gov/doc1.htm*). Both documents are available by request from NICEATM at the address provided above. Information submitted in response to this request will be incorporated into the background material provided to the Panel. The Panel's peer review meeting is anticipated to take place in early to mid-summer, and meeting information (including date and location) and public availability of the background document will be announced in a future **Federal Register** notice and will be posted on the ICCVAM website (*http://iccvam.niehs.nih.gov*). Information about studies with UDP should be sent to Dr. Stokes (contact information provided above).

Persons requesting additional information regarding the rationale for the OECD proposal to delete the OECD Guideline 401 can contact William T. Meyer, U.S. Environmental Protection Agency, Office of Pesticide Programs, phone: 703–305–7188; fax: 703–308–1805; e-mail: *Meyer.WilliamT@epa.gov.* Mail address: Ariel Rios Bldg., 1200 Pennsylvania Avenue, NW, Mail Code 7506C, Washington, DC 20460; Federal Express address: 1921 Jefferson Davis Highway, Room 1104H, Arlington, VA 22202.

Background Information

ICCVAM, with participation by 14 Federal regulatory and research agencies, was established in 1997 to coordinate cross-agency issues relating

to validation, acceptance, and national/international harmonization of toxicological test methods. ICCVAM seeks to promote the scientific validation and regulatory acceptance of toxicological test methods that will enhance the agencies' ability to assess risks and make decisions and that will refine, reduce, and replace animal use whenever possible. NICEATM provides administrative and technical support for ICCVAM and serves as a communication and information resource. NICEATM and ICCVAM collaborate to carry out related activities needed to develop, validate, and achieve regulatory acceptance of new and improved test methods applicable to Federal agencies. These activities may include:

1. Test Method Workshops are convened as needed to evaluate the adequacy of current test methods for assessing specific toxicities, to identify areas in need of improved or new testing methods, and to identify research and validation efforts that may be needed to develop a new test method.

2. Expert Panel Meetings are typically convened to evaluate the validation status of a test method following the completion of initial development and pre-validation studies. An Expert Panel is asked to recommend additional validation studies that might be helpful in further characterizing the usefulness of a method and to identify any additional research and development efforts that might support or enhance the accuracy and efficiency of a method.

3. Independent Peer Review Panel Meetings are typically convened following the completion of comprehensive validation studies on a test method. Panels are asked to develop scientific consensus on the usefulness and limitations of test methods and to generate information for specific human health and/or ecological risk assessment purposes. Following the review of a test method, ICCVAM forwards recommendations on its usefulness to agencies for their consideration. Federal agencies then determine the regulatory acceptability of a method according to their mandates.

Additional information about ICCVAM and NICEATM can be found at the website: *http://iccvam.niehs.nih.gov.*

Dated: February 11, 2000.

Samuel H. Wilson,

Deputy Director, National Institute of Environmental Health Sciences.

[FR Doc. 00–4010 Filed 2–17–00; 8:45 am]

BILLING CODE 4140–01–M

DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT

[Docket No. FR–4566–N–02]

Notice of Proposed Information, Collection: Comment Request—Hope for Homeownership of Single Family Homes (HOPE 3)

AGENCY: Office of the Assistant Secretary for Community Planning and Development, HUD.

ACTION: Notice.

SUMMARY: The proposed information collection requirement described below will be submitted to the Office of Management and Budget (OMB) for review, as required by the Paperwork Reduction Act. The Department is soliciting public comments on the subject proposal.

DATES: *Comments Due Date:* April 18, 2000.

ADDRESSES: Interested persons are invited to submit comments regarding this proposal. Comments should refer to the proposal by name and/or OMB Control Number and should be sent to: Shelia E. Jones, Reports Liaison Officer, Department of Housing and Urban Development, 451 7th Street, SW, Room 7232, Washington, DC 20410.

FOR FURTHER INFORMATION CONTACT: Patricia Mason, (202) 708–0614, ext. 4588 (this is not a toll-free number) for copies of the proposed forms and other available documents:

SUPPLEMENTARY INFORMATION: The Department is submitting the proposed information collection to OMB for review, as required by the Paperwork Reduction Act of 1995 (44 U.S.C. Chapter 35, as amended).

This Notice is soliciting comments from members of the public and affecting agencies concerning the proposed collection of information to: (1) Evaluate whether the proposed collection of information is necessary for the proper performance of the functions of the agency, including whether the information will have practical utility; (2) Evaluate the accuracy of the agency's estimate of the burden of the proposed collection of information; (3) Enhance the quality, utility, and clarity of the information to be collected; and (4) Minimize the burden of the collection of information on those who are to respond; including through the use of appropriate automated collection techniques or other forms of information technology, e.g., permitting electronic submission of responses.

This Notice also lists the following information:

Title of Proposal: HOPE for Homeownership of Single Family Homes (HOPE 3).

OMB Control Number, if applicable: 2506–0128.

Description of the need for the information and proposed use: The Homeownership Opportunities for People Everywhere (HOPE 3) Program provides Federal grants to develop and implement homeownership programs for low income people. This information is needed to assist HUD monitor grantees previously awarded HOPE 3 Program Implementation Grants through the collection of data in the Program's Cash and Management Information System, environmental review assessments and annual performance report requirements. The Department does not anticipate additional awards for the HOPE 3 Program.

Agency form numbers, if applicable: SF 424, HUD–40086, 40102–A, 40101–B, 40103, 40104, and 40105.

Members of affected public: State and local governments, nonprofit organizations.

Estimation of the total numbers of hours needed to prepare the information collection, including number of respondents, frequency of response, and hours of response: The Department estimates that the 158 respondents will require 15,490 hours annually (approximately 100 per respondent) to prepare the information collection.

Status of the proposed information collection: Reinstatement, with change, of a previously approved collection for which approval has expired.

Authority: Section 3506 of the Paperwork Reduction Act of 1995, 44 U.S.C. Chapter 35, as amended.

Dated: February 12, 2000.

Cardell Cooper,

Assistant Secretary for Community Planning and Development.

[FR Doc. 00–3879 Filed 2–17–00; 8:45 am]

BILLING CODE 4210–29–M

DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT

[Docket No. FR–4566–N–01]

Notice of Proposed Information Collection: Comment Request—Rural Housing and Economic Development

AGENCY: Office of the Assistant Secretary for Community Planning and Development, HUD.

ACTION: Notice.

SUMMARY: The proposed information collection requirement described below has been submitted to the Office of

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service

National Institute of Environmental Health Sciences (NIEHS), National Institutes of Health (NIH), National Toxicology Program (NTP); Notice of Peer Review Meeting on the Revised Up-and-Down Procedure (UDP) as an Alternative Test Method for Assessing Acute Oral Toxicity; Request for Comments

Summary

Pursuant to Public Law 103–43, notice is hereby given of a public meeting coordinated by the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) and the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and sponsored by NIEHS and the NTP. The agenda topic is the scientific peer review of the revised Up-and-Down Procedure, a method proposed as a replacement for the existing LD50 test for evaluating the acute oral toxicity potential of chemicals. The meeting will take place on July 25, 2000, from 8:30 a.m. to 5:30 p.m. at the Sheraton Crystal City Hotel, 1800 Jefferson Davis Highway, Arlington, VA 22202. The meeting is open to the public.

Background

ICCVAM, with participation by 14 Federal regulatory and research agencies and programs, was established in 1997 to coordinate issues relating to the development, validation, acceptance, and national/international harmonization of toxicological test methods. ICCVAM seeks to promote the scientific validation and regulatory acceptance of new and improved test methods applicable to Federal agencies including methods that may reduce and replace animal use, or that refine animal use to reduce or eliminate pain and distress. The Committee's functions include the coordination of interagency reviews of toxicological test methods and communication with stakeholders throughout the process of test method development and validation. The following Federal regulatory and research agencies and organizations participate in this effort:

Consumer Product Safety Commission
Department of Defense
Department of Energy
Department of Health and Human Services
  Agency for Toxic Substances and Disease Registry
  Food and Drug Administration
  National Institute for Occupational Safety and Health/CDC
  National Institutes of Health
  National Cancer Institute
  National Institute of Environmental Health Sciences
  National Library of Medicine
Department of the Interior
Department of Labor
  Occupational Safety and Health Administration
Department of Transportation
  Research and Special Programs Administration
Environmental Protection Agency

The NTP Center for the Evaluation of Alternative Toxicological Methods (NICEATM) was established in 1998 and provides operational support for the ICCVAM. NICEATM and ICCVAM seek to promote the validation and regulatory acceptance of new test methods that will enhance agencies' abilities to assess risks, and that will refine, reduce, and replace animal use. NICEATM and ICCVAM collaborate to carry out activities associated with the development, validation, and regulatory acceptance of proposed new and improved test methods. These activities may include:

Independent Peer Review Panel Meetings, which are typically convened following the completion of comprehensive validation studies on a test method. Independent peer review has been determined to be an essential prerequisite for consideration of a test method for regulatory acceptance. Peer Review Panels are asked to develop scientific consensus on the usefulness and limitations of test methods to generate information for specific human health and/or ecological risk assessment purposes. Following the independent peer review of a test method, ICCVAM forwards recommendations on their usefulness to agencies for their consideration. Federal agencies then determine the regulatory acceptability of a method according to their mandates.

Expert Panel Meetings, which are typically convened to evaluate the validation status of a method following the completion of initial development and pre-validation studies. An Expert Panel is asked to recommend additional validation studies that might be helpful in further characterizing the usefulness of a method and to identify any additional research and development efforts that might enhance the effectiveness of a method.

Test Method Workshops, which are convened, as needed, to evaluate the adequacy of current methods for assessing specific toxicities, to identify areas in need of improved or new testing methods, to identify research efforts that may be needed to develop new test methods, and to identify appropriate development and validation activities for proposed new methods.

Agenda

The agenda topic is the scientific peer review evaluation of the validation status of the revised Up-and-Down Procedure (UDP). This procedure is an updated version of the Organization for Economic Cooperation and Development (OECD) Test Guideline 425 (OECD Guideline for the Testing of Chemicals, Acute Oral Toxicity: Up-and-Down Procedure. Guideline 425, adopted September 21, 1998, OECD, Paris, France, *http://www.oecd.org/ehs/test*). The revised UDP is proposed as a substitute for the existing OECD Test Guideline 401 (OECD Guideline for the Testing of Chemicals, Acute Oral Toxicity, Guideline 401, adopted February 24, 1987, OECD, Paris, France). OECD has proposed that Guideline 401 should be deleted since three alternative methods are now available [OECD Document ENV/JM (99) 19, Test Guidelines Programme, Acute Oral Toxicity Testing: Data Needs and Animal Welfare Considerations, 29th Joint Meeting, June 8–11, 1999, Paris, France]. Prior to deletion of Guideline 401, U.S. agencies have requested that ICCVAM conduct an independent peer review of the revised UDP to determine the validity of the method as a substitute for Guideline 401. An Independent Peer Review Panel will (1) evaluate the extent to which established validation and acceptance criteria (''Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the ad hoc Interagency Coordinating Committee on the Validation of Alternative Methods,'' NIH Publication No. 97–3981, *http://ntpserver.niehs.nih.gov/htdocs/ICCVAM/iccvam.html*) have been addressed, and (2) will provide conclusions and recommendations regarding the usefulness and limitations of the method as a substitute for the traditional acute oral toxicity test method (OECD Guideline 401, 1987). The UDP has the potential to reduce the number of animals required to classify chemicals for acute oral toxicity compared to Guideline 401. A request for nominations of expert scientists for the Panel was previously published (FR 65, 8385–8386, February 18, 2000).

The meeting will begin at 8:30 a.m. on July 25 and will conclude by 5 p.m. There will be a brief orientation on ICCVAM and the ICCVAM review process, followed by a peer review of the revised UDP and supporting

information. The Peer Review Panel will discuss the usefulness of the UDP as an alternative to the traditional LD50 methods currently accepted by government regulatory authorities for the assessment of acute oral toxicity potential of chemicals.

Background Document Available for Comment

NICEATM has prepared a Background Review Document that includes the revised UDP protocol and documents supporting the basis and validity of the test method. Copies of the Up-and-Down Procedure Background Review Document and supporting documentation may be obtained from NICEATM, MD EC–17, P.O. Box 12233, Research Triangle Park, NC, 27709, Phone: 919–541–3398, Fax: 919–541–0947, E-mail: *ICCVAM@niehs.nih.gov*. A copy of the Background Review Document and comments submitted will be available for viewing Monday through Friday, from 12 noon to 4 p.m. EST at the U. S. Environmental Protection Agency, Office of Prevention, Pesticides and Toxic Substances, Non-Confidential Information Center, Room 607B, Northeast Mall, 401 M Street, SW, Washington, DC 20460. Thirty days prior to the meeting, a detailed agenda will be available on the web at: *http://iccvam.niehs.nih.gov* or by contacting NICEATM.

Persons requesting additional information regarding the rationale for the OECD proposal to delete the OECD Guideline 401 can contact William T. Meyer, U.S. Environmental Protection Agency, Office of Pesticide Programs, Phone: 703–305–7188; E-mail: *Meyer,WilliamT@epa.gov.* Mail address: Ariel Rios Bldg., 1200 Pennsylvania Ave., NW, Mail Code 7506C, Washington, DC 20460; Federal Express address: 1921 Jefferson Davis Highway, Room 1104H, Arlington, VA 22202.

Request for Comments

NICEATM invites the submission of written comments on the revised Up-and-Down Procedure, and submission of other available information and data on the UDP, including information about completed, ongoing, or planned studies. Written comments and additional information should include name, affiliation, mailing address, phone, fax, e-mail and sponsoring organization (if any), and should be sent by mail, fax, or e-mail to NICEATM at the address listed above. Comments may be submitted anytime before the meeting; however, comments should be submitted by June 15 in order to ensure time for adequate review by the Panel. Written comments will be made available to the Peer Review Panel members, ICCVAM agency representatives and experts, and attendees at the meeting and will be included in the resource materials assembled on the UDP.

The Expert Panel Meeting will be open to the public, and time will be provided for presentation of public oral comments at designated times during the meeting. Speakers will be assigned on a first-come, first-serve basis and up to seven minutes will be allotted to each speaker. In order to facilitate planning, members of the public who wish to present oral statements at the meeting should contact NICEATM as soon as possible, but no later than July 18, 2000. Persons registering to make comments are asked to provide, if possible, a written copy of their statement in advance so that copies can be made and distributed to the Peer Review Panel members for their timely consideration prior to the meeting. Written statements can supplement and expand the oral presentation, and each speaker is asked to provide his/her name, affiliation, mailing address, phone, fax, e-mail and supporting organization (if any). Registration for making public comments will also be available on-site. If registering on-site to speak and reading oral comments from printed copy, the speaker is asked to bring 50 copies of the text. These copies will be distributed to the Panel and supplement the record.

Summary minutes from the meeting and the final report from the Peer Review Panel will be prepared and made available upon request to NICEATM (address provided above). These documents will also be made available via the internet at the website: *http://iccvam.niehs.nih.gov.*

Additional information about ICCVAM and NICEATM can be found at the website:*http://iccvam.niehs.nih.gov.*

Dated: May 22, 2000.

Samuel H. Wilson,

Deputy Director, National Institute of Environmental Health Sciences.

[FR Doc. 00–13734 Filed 5–31–00; 8:45 am]

BILLING CODE 4140–01–P

DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT

[Docket No. FR–4563–N–06]

Notice of Proposed Information Collection for Tenant Opportunities Semi-Annual Report

AGENCY: Office of the Assistant Secretary for Public and Indian Housing, HUD.

ACTION: Notice.

SUMMARY: The proposed information collection requirement described below will be submitted to the Office of Management and Budget (OMB) for review, as required by the Paperwork Reduction Act. The Department is soliciting public comments on the subject proposal.

DATES: *Comments Due Date:* July 31, 2000.

ADDRESSES: Interested Persons are invited to submit comments regarding this proposal. Comments should refer to the proposal by name and/or OMB Control number and should be sent to: Mildred M. Hamman, Reports Liaison Officer, Public and Indian Housing, Department of Housing and Urban Development, 451 7th Street, SW, Room 4238, Washington, DC 20410–5000.

FOR FURTHER INFORMATION CONTACT: Mildred M. Hamman, (202) 708–3642, extension 4128, for copies of the proposed forms and other available documents. (This is not a toll-free number).

SUPPLEMENTARY INFORMATION: The Department will submit the proposed information collection to OMB for review, as required by the Paperwork Reduction Act of 1995 (44 U.S.C. Chapter 35, as amended).

This Notice is soliciting comments from members of the public and affected agencies concerning the proposed collection of information to: (1) Evaluate whether the proposed collection of information is necessary for the proper performance of the functions of the agency, including whether the information will have practical utility; (2) evaluate the accuracy of the agency's estimate of the burden of the proposed collection of information; (3) enhance the quality, utility, and clarity of the information to be collected; and (4) minimize the burden of the collection of information on those who are to respond, including through the use of appropriate automated collection techniques or other forms of information technology; *e.g.*, permitting electronic submission of responses.

This Notice also lists the following information:

Title of Proposal: Tenant Opportunities Semi-Annual Report.

OMB Control Number: 2577–0087.

Description of the need for the information and proposed use: Grantees participating in TOP are required to submit Semiannual Report (Form HUD–52370), which will evaluate the progress in carrying out the approved TOP workplan/budget. Grantees shall submit the report on a semiannual basis for the

DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

National Institute of Mental Health; Notice of Closed Meeting

Pursuant to section 10(d) of the Federal Advisory Committee Act, as amended (5 U.S.C. Appendix 2), notice is hereby given of the following meeting.

The meeting will be closed to the public in accordance with the provisions set forth in sections 552b(c)(4) and 552b(c)(6), Title 5 U.S.C., as amended. The contract proposals and the discussions could disclose confidential trade secrets or commercial property such as patentable material, and personal information concerning individuals associated with the contract proposals, the disclosure of which would constitute a clearly unwarranted invasion of personal privacy.

Name of Committee: National Institute of Mental Health Special Emphasis Panel.

Date: July 10, 2001.

Time: 9 a.m. to 12 p.m.

Agenda: To review and evaluate contract proposals.

Place: Marriott Suites Bethesda, 6711 Democracy Boulevard, Bethesda, MD 20817.

Contact Person: Peter J. Sheridan, Scientific Review Administrator, Division of Extramural Activities, National Institute of mental Health, NIH, Neuroscience Center, 6001 Executive Blvd., Room 6142, MSC 9606, Bethesda, MD 20892–9606, 301–443–1513, psherida@mail.nih.gov

(Catalogue of Federal Domestic Assistance Program Nos. 93.242, Mental Health Research Grants; 93.281, Scientist Development Award, Scientist Development Award for Clinicians, and Research Scientist Award; 93.282, Mental Health National Research Service Awards for Research Training, National Institutes of Health, HHS)

Dated: June 18, 2001.

LaVerne Y. Stringfield,

Director, Office of Federal Advisory Committee Policy.

[FR Doc. 01–15769 Filed 6–21–01; 8:45 am]

BILLING CODE 4140–01–M

DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

Center for Scientific Review; Notice of Closed Meeting

Pursuant to section 10(d) of the Federal Advisory Committee Act, as amended (5 U.S.C. Appendix 2), notice is hereby given of the following meeting.

The meeting will be closed to the public in accordance with the provisions set forth in sections 552(b)(c)(4) and 552b(c)(6), Title 5 U.S.C., as amended. The grant applications and the discussions could disclose confidential trade secrets or commercial property such as patentable material, and personal information concerning individuals associated with the grant applications, the disclosure of which would constitute a clearly unwarranted invasion of personal privacy.

Name of Committee: Center for Scientific Review Special Emphasis Panel.

Date: June 20, 2001.

Time: 2:30 p.m. to 3:30 p.m.

Agenda: To review and evaluate grant applications.

Place: NIH, Rockledge 2, Bethesda, MD 20892 (Telephone Conference Call).

Contact Person: Julian L. Azorlosa, Scientific Review Administrator, Center for Scientific Review, National Institutes of Health, 6701 Rockledge Drive, Room 3190, MSC 7848, Bethesda, MD 20892, (301) 435–1507.

This notice is being published less than 15 days prior to the meeting due to the timing limitations imposed by the review and funding cycle.

(Catalogue of Federal Domestic Assistance Program Nos. 93.306, Comparative Medicine, 93.306; 93.333, Clinical Research, 93.333, 93.337, 93.393–93.396, 93.837–93.844, 93.846–93.878, 93.892, 93.893, National Institutes of Health, HHS)

Dated: June 18, 2001.

LaVerne Y. Stringfield,

Director, Office of Federal Advisory Committee Policy.

[FR Doc. 01–15765 Filed 6–21–01; 8:45 am]

BILLING CODE 4140–01–M

DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

Center for Scientific Review; Amended Notice of Meeting

Notice is hereby given of a change in the meeting of the Center for Scientific Review Special Emphasis Panel, June 21, 2001, 8:30 a.m. to June 22, 2001, 6 p.m., River Inn, 924 25th Street, NW., Washington, DC, 20037 which was published in the **Federal Register** on June 12, 2001, 66 FR 31683–31685.

The meeting will be one day only June 21, 2001. The time and location remain the same. The meeting is closed to the public.

Dated: June 18, 2001.

LaVerne Y. Stringfield,

Director, Office of Federal Advisory Committee Policy.

[FR Doc. 01–15768 Filed 6–21–01; 8:45 am]

BILLING CODE 4140–01–M

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service

National Institute of Environmental Health Sciences (NIEHS); National Toxicology Program (NTP); The Revised Draft Up-and-Down Procedure for Assessing Acute Oral Toxicity: Notice of Availability and Request for Public Comments

Summary

Notice is hereby given of the availability of a revised draft Up-and-Down Procedure for assessing acute oral toxicity and solicitation of public comment. Documents available include: (1) A revised draft Up-and-Down Procedure (UDP) test guideline (hereafter, revised draft UDP); (2) A procedure incorporated into the revised draft UDP for calculating the confidence interval for the estimated median lethal dose (LD50); and (3) A software program for use in establishing test doses, determining when to stop the test, and estimating the LD50 and the confidence interval for the estimated LD50.

Availability of Revised Draft UDP Documents

The revised draft UDP was proposed by the U.S. Environmental Protection Agency (U.S. EPA) to the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) as an alternate for the existing conventional LD50 test (EPA 870.1100) used to evaluate the acute oral toxicity of chemicals. A previous version of the draft UDP was reviewed by the UDP Peer Review Panel (hereafter, Panel) at a meeting on July 25, 2000 organized by the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and ICCVAM. This revised draft UDP incorporates modifications made in response to the conclusions and recommendations of the Panel and may be obtained electronically from the NICEATM/ICCVAM web site at http:/iccvam.niehs.nih.gov/methods/udpdocs/udprpt/udp_ciprop.htm. For a paper copy (a limited number are available), contact NICEATM at (919) 541–3398, or via e-mail at niceatm@niehs.nih.gov.

The proposed procedure for calculating the confidence interval for the estimated LD50 is a statistical calculation and does not require the use of test animals beyond what is needed to estimate the LD50. This procedure helps to place the estimated LD50 in a statistical context for hazard and risk assessment purposes. The confidence

interval procedure may be obtained electronically from the NICEATM/ICCVAM web site at http://iccvam.niehs.nih.gov/methods/udpdocs/udprpt/udp_ciprop.htm. For a paper copy (a limited number are available), contact NICEATM at (919) 541–3398, or via e-mail at niceatm@niehs.nih.gov. For technical clarification or questions regarding the confidence interval procedure, contact Dr. Amy Rispin, U.S. EPA, by telephone at (703) 305–5989 or via e-mail at rispin.amy@epa.gov.

Because the generation of parameters for this revised draft UDP is computationally intensive, the U.S. EPA developed a simple-to-use software program to aid in dose selection, test-stopping decisions, calculation of an estimate of the LD50, and calculation of a confidence interval around the LD50. The confidence interval procedure may be obtained electronically from the NICEATM/ICCVAM web site at http://iccvam.niehs.nih.gov/methods/udpdocs/udprpt/udp_ciprop.htm. To obtain a diskette of this software program, (a limited number are available), contact NICEATM at (919) 541–3398 or via e-mail at niceatm@niehs.nih.gov. For technical clarification or questions regarding the software package contact Dr. Elizabeth Margosches, U.S. EPA, by telephone at (202) 260–1511 or via e-mail at margosches.elizabeth@epa.gov, or Ms. Deborah McCall, U.S. EPA, by telephone at (703) 305–7109, or via e-mail at mccall.deborah@epa.gov.

Request for Public Comment

NICEATM invites written public comments on the revised draft UDP, the confidence interval proposal, and the software program. Comments should be sent to NICEATM through August 6, 2001. Comments submitted via e-mail are preferred; the acceptable file formats are MS Word (Office 98 or older), plain text, or PDF. Comments should be sent to Dr. William S. Stokes, Director, NICEATM, NIEHS, MD EC–17, P.O. Box 12233, Research Triangle Park, NC, 27709; telephone 919–541–2384; fax 919–541–0947; e-mail niceatm@niehs.nih.gov. Persons submitting written comments should include their contact information (name, affiliation, address, telephone and fax numbers, and e-mail) and sponsoring organization, if any. Public comments received in response to this **Federal Register** notice will be posted on the NICEATM/ICCVAM web site (http://iccvam.niehs.nih.gov). In addition, they will be available for viewing Monday through Friday, from noon to 4 p.m., excluding legal holidays, at the U.S. EPA under docket control number: AR–228, Up-and-Down Procedure. [U.S. EPA, Office of Prevention, Pesticides, and Toxic Substances, Non-Confidential Information Center, Room 607B, Northeast Mall, 401 M Street, SW., Washington, DC 20460, telephone: (202) 260–7099]. This docket also contains background and supporting materials for the revised draft UDP.

The comments will also be provided to the Panel for consideration in preparation for a final meeting tentatively planned for August 2001. This meeting is anticipated to be held as a teleconference with opportunity for public participation. An announcement of the Panel meeting with additional details will be published in a future **Federal Register** notice. The focus of this meeting will be to discuss the revised draft UDP, the proposed procedure for calculating the confidence interval for the estimated LD50, and the software program. Following the Panel meeting, a final report of the Panel's findings and recommendations will be published and made available to the public through NICEATM. In accordance with Public Law 106–545, ICCVAM will develop and forward test recommendations on the UDP to Federal agencies for their consideration. The ICCVAM recommendations will also be made available to the public.

Background

In 1999, the Organization for Economic Cooperation and Development (OECD) proposed deletion of its standard test guideline (TG) for assessing the acute oral toxicity of chemicals (TG 401; OECD, 1987). The rationale for deletion was that three alternative acute toxicity test methods had previously been adopted and could be used instead. Each method uses fewer animals than the procedure described in TG 401. One of these test methods is the UDP (OECD TG 425). Prior to formal deletion of TG 401, OECD determined that it was necessary to revise the three alternative methods to conform to the newly harmonized OECD hazard classification scheme (OECD, 1998). The U.S. EPA agreed to organize a Technical Task Force to revise the UDP (OECD TG 425). The revised UDP test method included two procedures different from the original UDP: a Limit Test for substances anticipated having minimal toxicity, and a Supplemental Test to determine the slope and confidence interval for the dose-response curve.

ICCVAM and NICEATM convened an international independent scientific peer review panel July 25, 2000, to evaluate the validation status of the revised UDP. The Panel concluded that the revised UDP Primary Test provided an improved estimate of acute oral toxicity with a reduction in the number of animals used compared to the existing conventional LD50 test (e.g., EPA 870.1100, TG 401). The Panel concluded that the proposed Limit Test procedure would be expected to perform as well as or better than the currently used EPA 870.1100 or TG 401 limit test for hazard classification, while using fewer animals. The Panel did not recommend the proposed UDP Supplemental Test procedure for use. Information on previous deliberations of the Panel can be found on the Internet at http://iccvam.niehs.nih.gov/udp.htm.

In recognition of the need for a procedure to calculate the confidence interval for the estimated median lethal dose determined using the UDP, the UDP Technical Task Force developed a procedure for use with UDP data from the primary procedure. As recommended by the Panel, the Supplemental Procedure has been deleted in the revised draft UDP and no further work on a procedure to generate dose-response slope information has been proposed. A specialized software program was subsequently developed by the U.S. EPA to facilitate implementation and use of the revised UDP.

Background for the UDP, including the availability of review materials, can be found in previous **Federal Register** notices (see FR Volume 65, Number 34, pages 8385–8386, February 18, 2000, and FR Volume 65, Number 106, pages 35109–35110, June 1, 2000). Minutes from the UDP Peer Review Panel meeting held July 25, 2000, may be found at http://iccvam.niehs.nih.gov/udp.htm.

Additional Information About ICCVAM and NICEATM

ICCVAM, with 15 participating Federal agencies, was established in 1997 to coordinate interagency issues on toxicological test method development, validation, regulatory acceptance, and national and international harmonization. The ICCVAM Authorization Act of 2000 (Pub. L. 106–545) formally authorized and designated ICCVAM as a permanent committee. The NICEATM was established in 1998 to collaborate with the ICCVAM to facilitate the development, scientific review, and validation of novel toxicological methods that predict human health risks while reducing, refining, and/or replacing animal tests and to promote communication with stakeholders. The NICEATM is located at the NIEHS in Research Triangle Park, NC. Additional information concerning ICCVAM and NICEATM can be found

on the ICCVAM/NICEATM web site at http://iccvam.niehs.nih.gov.

References

U.S. EPA (1998). Health Effects Test Guidelines, OPPTS 870.1100, Acute Oral Toxicity. Washington, DC: U.S. Environmental Protection Agency, 1998. Available on the Internet at http://www.epa.gov/docs/OPPTS_Harmonized/870_Health_Effects_Test_Guidelines/Series/.

OECD (1987). TG 401. OECD Guideline for the Testing of Chemicals, Acute Oral Toxicity, Adopted February 24, 1987, OECD, Paris, France.

OECD (1998). Harmonized Integrated Hazard Classification System for Human Health and Environmental Effects of Chemical Substances as endorsed by the 28th Joint Meeting of the Chemicals Committee and Working Party on Chemicals in November 1998. Available on the Internet at http://www.oecd.org/ehs/Class/HCL6.htm.

Dated: June 6, 2001.

Samuel H. Wilson,

Deputy Director, National Institute of Environmental Health Sciences.

[FR Doc. 01–15770 Filed 6–21–01; 8:45 am]

BILLING CODE 4140–01–U

DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT

[Docket No. FR–4655–N–16]

Notice of Proposed Information Collection: Comment Request; Congregate Housing Services Program (CHSP)

AGENCY: Office of the Assistant Secretary for Housing, HUD.

ACTION: Notice.

SUMMARY: The proposed information collection requirement described below will be submitted to the Office of Management and Budget (OMB) for review, as required by the Paperwork Reduction Act. The department is soliciting public comments on the subject proposal.

DATES: *Comments Due Date:* August 21, 2001.

ADDRESSES: Interested persons are invited to submit comments regarding this proposal. Comments should refer to the proposal by name and/or OMB Control Number and should be sent to: Wayne Eddins, Reports Management Officer, Department of Housing and Urban Development, 451 7th Street, SW, L'Enfant Building, Room 8202, Washington, D.C. 20410.

FOR FURTHER INFORMATION CONTACT: Carissa Janis, Office of Housing Assistance and Grants Management, U.S. Department of Housing and Urban Development, 451 7th Street, SW, Washington, DC 20410, telephone number (202) 708–2866, extension 2487 (this is not a toll-free number), for copies of the proposed forms and other available information.

SUPPLEMENTARY INFORMATION: The Department is submitting the proposed information collection to OMB for review, as required by the Paperwork Reduction Act of 1955 (44 U.S.C. Chapter 35, as amended).

This Notice is soliciting comments from members of the public and affected agencies concerning the proposed collection of information to: (1) Evaluate whether the proposed collection is necessary for the proper performance of the functions of the agency, including whether the information will have practical utility; (2) Evaluate the accuracy of the agency's estimate of the burden of the proposed collection of information; (3) Enhance the quality, utility, and clarity of the information to be collected; and (4) Minimize the burden of the collection of information on those who are to respond; including the use of appropriate automated collection techniques or other forms of information technology, e.g., permitting electronic submission of responses.

This Notice also lists the following information:

Title of Proposal: Congregate Housing Services Program (CHSP).

OMB Control Number, if applicable: 2502–0485.

Description of the need for the information and proposed use: Completion of the Annual Report by grantees provides HUD with essential information about who the grant is serving and what sort of services the individual receive through the use of grant funds. The Summary Budget is a matrix of budgeted yearly costs, which shows the services funded through the grant and demonstrates how matching funds, participants fees, and grant funds will be used in tandem to operate the grant program. Field staff approve this annual budget and request annual extension funds according to the budget. Field staff can also determine if grantees are meeting statutory and regulatory requirements through the evaluation of this budget. HUD will use the Payment Voucher to monitor the use of grant funds for eligible activities over the term of the grant. The Grantee may similarly use the Payment Voucher to track and record their request for payment reimbursement for grant-funded activities over the term of the grant. The grantee may similarly use the Payment Voucher to track and record their request for payment reimbursement for grant-funded activities.

Agency from numbers, if applicable: HUD–90006, HUD–90198, HUD–91180–A.

Estimation of the total number of hours needed to prepare the information collection including number of respondents, frequency of response, and hours of response: The estimated number of respondents is 81, the frequency of responses is annually, estimated time to compete is approximately 4 hours for HUD–90006; .25 hours for HUD–90198; 3.5 hours for HUD–91180–A; and 2 hours for SF–269, and the total annual burden hours requested for this collection is 1,013.

Status of the proposed information collection: Reinstatement with change, of previously approved collection for which approval has expired.

Authority: The Paperwork Reduction Act of 1995, 44 U.S.C., Chapter 35, as amended.

Dated: June 1, 2001.

Sean G. Cassidy,

General Deputy, Assistant Secretary for Housing—Deputy Federal Housing Commissioner.

[FR Doc. 01–15685 Filed 6–21–01; 8:45 am]

BILLING CODE 4210–27–M

DEPARTMENT OF HOUSING AND URBAN DEVELOPMENT

[Docket No. FR–4650–N–4]

Notice of Submission of Proposed Information Collection to OMB; Public Housing Assessment System (PHAS) Memorandum of Agreement (MOA) and Improvement Plan (IP)

AGENCY: Office of the Chief Information Officer, HUD.

ACTION: Notice.

SUMMARY: The proposed information collection requirement described below has been submitted to the Office of Management and Budget (OMB) for review, as required by the Paperwork Reduction Act. The Department is soliciting public comments on the subject proposal.

DATES: *Comments Due Date:* July 23, 2001.

ADDRESSES: Interested persons are invited to submit comments regarding this proposal. Comment should refer to the proposal by name and/or OMB approval number and should be sent to: Joseph F. Lackey, Jr., OMB Desk Officer, Office of Management and Budget, Room 10235, New Executive Office Building, Washington, DC 20503.

DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

National Institute of Child Health and Human Development; Notice of Closed Meeting

Pursuant to section 10(d) of the Federal Advisory Committee Act, as amended (5 U.S.C. Appendix 2), notice is hereby given of the following meeting.

The meeting will be closed to the public in accordance with the provisions set forth in sections 552b(c)(4) and 552b(c)(6), Title 5 U.S.C., as amended. The contract proposals and the discussions could disclose confidential trade secrets or commercial property such as patentable material, and personal information concerning individuals associated with the contract proposals, the disclosure of which would constitute a clearly unwarranted invasion of personal privacy.

Name of Committee: National Institute of Child Health and Human Development Special Emphasis Panel.

Date: August 6–7, 2001.

Time: 8 am to 12 pm.

Agenda: To review and evaluate contract proposals.

Place: Double Tree Hotel, 1750 Rockville Pike, Rockville, MD 20852.

Contact Person: Hameed Khan, PhD, Scientific Review Administrator, Division of Scientific Review, National Institute of Child Health and Human Development, National Institutes of Health, 6100 Executive Blvd., Room 5E01, Bethesda, MD 20892, (301) 496–1485.

(Catalogue of Federal Domestic Assistance Program Nos. 93.209, Contraception and Infertility Loan Repayment Program; 93.864, Population Research; 93.865, Research for Mothers and Children; 93.929, Center for Medical Rehabilitation Research, National Institutes of Health, HHS)

Dated: June 29, 2001.

LaVerne Y. Stringfield,

Director, Office of Federal Advisory Committee Policy.

[FR Doc. 01–17283 Filed 7–10–01; 8:45 am]

BILLING CODE 4140–01–M

DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

National Institute of Allergy and Infectious Diseases; Notice of Closed Meeting

Pursuant to seciton 10(d) of the Federal Advisory Committee Act, as amended (4 U.S.C. Appendix 2), notice is hereby given of the following meeting.

The meeting will be closed to the public in accordance with the provisions set forth in sections 552b(c)(4) and 552b(c)(6), Title 5 U.S.C., as amended. The grant applications and the discussions could disclose confidential trade secrets or commercial property such as patentable material, and personal information concerning individuals associated with the grant applications, the disclosure of which would constitute a clearly unwarranted invasion of personal privacy.

Name of Committee: National Institute of Allergy and Infectious Diseases Special Emphasis Panel.

Date: July 12, 2001.

Time: 10 am to 10:30 am.

Agenda: To review and evaluate grant applications.

Place: 6700 B Rockledge Drive, Bethesda, MD 20892 (Telephone Conference Call).

Contact Person: Roberta Binder, PhD, Scientific Review Administrator, Division of Extramural Activities, NIAID, 6700B Rockledge Drive, Rm 2155, Bethesda, MD 20892, 301–496–7966, rb169n@nih.gov.

This notice is being published less than 15 days prior to the meeting due to the timing limitations imposed by the review and funding cycle.

(Catalogue of Federal Domestic Assistance Program Nos. 93.855, Allergy, Immunology, and Transplantation Research; 93.856, Microbiology and Infectious Diseases Research, National Institutes of Health, HHS)

Dated: June 29, 2001.

LaVerne Y. Stringfield,

Director, Office of Federal Advisory Committee Policy.

[FR Doc. 01–17287 Filed 7–10–01; 8:45 am]

BILLING CODE 4140–01–M

DEPARTMENT OF HEALTH AND HUMAN SERVICES

National Institutes of Health

National Library of Medicine; Notice of Closed Meeting

Pursuant to section 10(d) of the Federal Advisory Committee Act, as amended (5 U.S.C. Appendix 2), notice is hereby given of the following meeting.

The meeting will be closed to the public in accordance with the provisions set forth in sections 552b(c)(4) and 552b(c)(6), Title 5 U.S.C., as amended. The contract proposals and the discussions could disclose confidential trade secrets or commercial property such as patentable material, and personal information concerning individuals associated with the contract proposals, the disclosure of which would constitute a clearly unwarranted invasion of personal privacy.

Name of Committee: National Library of Medicine Special Emphasis Panel.

Date: July 5, 2001.

Time: 2 pm to 3 pm.

Agenda: To review and evaluate contract proposals.

Place: National Library of Medicine, Building 38A, HPCC Conference Room B1N30Q, 8600 Rockville Pike, Bethesda, MD 20894, (Telephone Conference Call).

Contact Person: Merlyn M. Rodrigues, Medical Officer/SRA, National Library of Medicine, Extramural Programs, 6705 Rockledge Drive, Suite 301, Bethesda, MD 20894.

This notice is being published less than 15 days prior to the meeting due to the timing limitations imposed by the review and funding cycle.

(Catalogue of Federal Domestic Assistance Program Nos. 93.879, Medical Library Assistance, National Institutes of Health, HHS)

Dated: June 29, 2001.

LaVerne Y. Stringfield,

Director, Office of Federal Advisory Committee Policy.

[FR Doc. 01–17276 Filed 7–10–01; 8:45 am]

BILLING CODE 4140–01–M

DEPARTMENT OF HEALTH AND HUMAN SERVICES

Public Health Service

National Institute of Environmental Health Sciences (NIEHS); National Toxicology Program (NTP); Peer Review Panel for the Up-and-Down Procedure (UDP): Notice of Meeting

Summary

Pursuant to section 10(d) of the Federal Advisory Committee Act, as amended (5 U.S.C. Appendix 2), notice is hereby given of a public teleconference meeting of the Up-and-Down Procedure (UDP) independent scientific peer review panel (Panel). The teleconference is scheduled for Tuesday, August 21, 2001, from 10:00 a.m.–12:00 p.m. EDT. The agenda for this meeting will focus on a discussion of the following: (1) The revised draft Up-and-Down Procedure (UDP), modified in response to recommendations from the July 2000 Panel meeting; (2) a proposed procedure for calculating the confidence interval for the estimated LD50; and (3) a software program to aid in dose selection, test-stopping decisions, calculation of an estimated LD50, and calculation of a confidence interval around the LD50.

Following the Panel meeting, a final report of the Panel's findings and recommendations will be published and made available to the public through the NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM). In accordance with Public Law 106–545, the Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM) will develop and forward test recommendations on the UDP to Federal agencies for their consideration. The ICCVAM recommendations will also be made available to the public through the NICEATM.

Background, including the availability of review materials, can be found in previous **Federal Register** notices (see FR Volume 66, Number 121, pages 33550–33552; FR Volume 65, Number 34, pages 8385–8386; and FR Volume 65, Number 106, pages 35109–35110). The **Federal Register** notice (Volume 66, Number 121) invites written public comments on the materials being discussed at the Panel meeting. Comments received by the August 6, 2001 deadline will be made available to the Panel prior to the August 21 teleconference.

Meeting information

Panel members will participate in the meeting via teleconference. The teleconference will originate from Room 3162, 3rd Floor, NIEHS, 79 T.W. Alexander Drive, Bldg. 4401, Research Triangle Park, NC and NICEATM staff will be on hand to coordinate the teleconference. The public is invited to attend with attendance limited only by the space available in Room 3162. To attend this meeting, please contact Ms. Loretta Frye, NICEATM, NIEHS, 79 Alexander Drive, Bldg. 4401, P.O. Box 12233, EC–17, Research Triangle Park, NC 27709; telephone (919) 541–3138; fax (919) 541–0947; or email *frye@niehs.nih.gov.* Arrangements to attend the meeting, including the need for special accommodation, (e.g., wheelchair access), should be made with the NIEHS/NICEATM staff by 12:00 noon EDT on Tuesday, August 14, 2001.

Request for Public Comment

While written public comments are requested and preferred, there will be an opportunity for oral public comments. For this teleconference meeting, oral comments by individual speakers will usually be limited to no more than three minutes per speaker. Persons registering to make oral comments are asked to provide their name, affiliation, mailing address, phone, fax, e-mail, and sponsoring organization. To facilitate planning for the meeting, persons interested in providing formal oral comments are asked to notify Ms. Loretta Frye (contact information provided above) in writing (email, fax, or mail) no later than 12:00 noon EDT on Tuesday, August 14, 2001. Persons registering to make oral comments are asked, if possible, to provide a copy of their statement to Ms. Loretta Frye by August 14, to enable review by the Panel and NICEATM staff prior to the meeting.

Dated: July 3, 2001.

Samuel H. Wilson,

Deputy Director, National Institute of Environmental Health Sciences.

[FR Doc. 01–17288 Filed 7–10–01; 8:45 am]

BILLING CODE 4140–01–P

DEPARTMENT OF THE INTERIOR

Fish and Wildlife Service

Notice of Receipt of Applications for Permit

Endangered Species

The following applicants have applied for a permit to conduct certain activities with endangered species. This notice is provided pursuant to Section 10(c) of the Endangered Species Act of 1973, as amended (16 U.S.C. 1531, *et seq.*). Written data or comments should be submitted to the Director, U.S. Fish and Wildlife Service, Division of Management Authority, 4401 North Fairfax Drive, Room 700, Arlington, Virginia 22203 and must be received by the Director within 30 days of the date of this publication.

Applicant: Dale Lee Nunez, Portland, OR, PRT–044912

The applicant requests a permit to import the sport-hunted trophy of one male bontebok (*Damaliscus pygargus dorcas*) culled from a captive herd maintained under the management program of the Republic of South Africa, for the purposes of enhancement of the survival of the species.

Applicant: Edward W. Berkeley, Portland, OR, PRT–044913

The applicant requests a permit to import the sport-hunted trophy of one male bontebok (*Damaliscus pygargus dorcas*) culled from a captive herd maintained under the management program of the Republic of South Africa, for the purposes of enhancement of the survival of the species.

Applicant: Bowmanville Zoo, Ontario, Canada, PRT–044983

The applicant requests a permit to import and re-export a captive-born jaguar (*Panthera onca*) and progeny of the animals currently held by the applicant and any animals acquired in the United States to/from worldwide locations to enhance the survival of the species through conservation education. This notification covers activities conducted by the applicant over a three-year period.

Applicant: Dr. M. F. Marcone, Department of Food Science, University of Guelph, Ontario, Canada, PRT–044611

The applicant requests a permit to import and re-export specimens of the endangered plants, *Achyranthes splendens* var. *rotundata* and *Nototrichium humile,* to and from various research facilities in the United States for the purposes of scientific research. This notification covers the activities conducted by the applicant over a five-year period.

Marine Mammals and Endangered Species

The public is invited to comment on the following application for a permit to conduct certain activities with endangered marine mammals. The application was submitted to satisfy requirements of the Marine Mammal Protection Act of 1972, *as amended* (16 U.S.C. 1361 *et seq.*), the Endangered Species Act of 1973, *as amended* (16 U.S.C. 1531, *et seq.*), and the regulations governing marine mammals (50 CFR 18) and endangered species (50 CFR 17).

Applicant: Harbor Branch Oceanographic Institution, Fort Pierce, FL PRT–038605.

Permit Type: Take for Scientific Research.

Name and Number of Animals: West Indian Manatee, *Trichechus manatus,* 8.

Summary of Activity to be Authorized: The applicant requests a permit to transfer 6 captive held, 2 captive born, as well as 1 Pre-Act, specimens, from Homosassa Springs Wildlife Park, Homosassa, FL, to their facility at Ft. Pierce, Florida, for the purpose of scientific research.

Source of Marine Mammals: Captive held and captive born.

Period of Activity: Up to 5 years if issued.

Concurrent with the publication of this notice in the **Federal Register**, the Division of Management Authority is forwarding copies of the above application to the Marine Mammal Commission and the Committee of Scientific Advisors for their review.

The public is invited to comment on the following application for a permit to conduct certain activities with marine mammals. The application was

APPENDIX E

Summary Minutes and Public Comments from the UDP Meetings

E–1 Minutes and Public Comments of the Peer Review Panel Meeting E-3
July 25, 2000 in Crystal City, Arlington, VA

E–2 Minutes of the Peer Review Panel Meeting E-13
August 21, 2001 in Research Triangle Park, NC

E–3 Submitted Public Comment for the August 21, 2001 Meeting E-19

Department of Health and Human Services
National Institutes of Health
National Institute of Environmental Health Sciences
Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)
Special Emphasis Panel

Minutes of the Peer Review Panel Meeting on the Revised Up-and-Down Procedure: A

Test Method for Assessing the Acute Oral Toxicity Potential of Chemicals

Introduction

A public meeting of an independent peer review panel was convened on July 25, 2000, at the Sheraton Imperial-Crystal City in Arlington, VA, to review the Revised Up-and-Down Procedure (UDP). The purpose of this meeting was to evaluate the validation status of the UDP as a replacement for the conventional LD50 test (OECD TG401; EPA OPPTS 870.1100). The meeting was organized by the Interagency Coordinating Committee for the Validation of Alternative Methods (ICCVAM) and the National Toxicology Program (NTP) Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM) and sponsored by the National Institute of Environmental Health Sciences (NIEHS) and the NTP. A comprehensive report of the peer review panel is provided as an attachment to these minutes.

The following expert scientists served on the peer review panel:

- Curtis D. Klaassen, Ph.D., D.A.B.T., D.A.T.S., Head, Section on Toxicology, Department of Pharmacology, Toxicology and Therapeutics, University of Kansas Medical Center, Kansas City, KS (Panel Co-Chair)
- Diane K. Gerken, D.V.M., Ph.D., D.A.B.T., D.A.B.V.T., Manager, Toxicology, Battelle, Columbus, OH (Panel Co-Chair)
- George Alexeeff, Ph.D., D.A.B.T., Deputy Director for Scientific Affairs, Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, Sacramento, CA.
- Bas J. Blaauboer, Ph.D., Associate Professor of Toxicology, Research Institute of Toxicology, Utrecht University, Utrecht, The Netherlands
- Kimberly Bonnette, M.S., LATG, Manager of Acute Toxicology, Springborn Laboratories, Inc., Spencerville, OH.
- Phil P.A. Botham, Ph.D., MRCPath, Section Head-Toxicity, Central Toxicology Laboratory, Zeneca, Ltd., Cheshire, United Kingdom
- Robert Condon, Ph.D., Consulting Biostatistician (Retired from the FDA Center for Veterinary Medicine), Myersville, MD
- Robert Copeland, Ph.D., Associate Professor, College of Medicine, Howard University, Washington, DC
- Wyman Dorough, Ph.D., Professor and Toxicologist, Mississippi State University, Starkville, MS

- Nancy Flournoy, Ph.D., Professor, Department of Mathematics and Statistics, American University, Washington, DC
- Charles Hastings, Ph.D., Manager of Toxicology, BASF Corporation, RTP, NC
- Wallace Hayes, Ph.D., D.A.B.T., D.A.T.S., Vice President for Corporate Product Integrity, The Gillette Company, Boston, MA
- Janice Kuhn, Ph.D., D.A.B.T, Group Leader, Toxicology, Stillmeadow, Inc., Sugar Land, TX
- John Reeve, M.S., National Manager (Toxicology and Residues), New Zealand Ministry of Agriculture and Forestry, Food Assurance Authority, ACVM Group, Wellington, New Zealand
- Robert Scala, Ph.D., D.A.B.T., D.A.T.S., Toxicology Consultant, Tucson, AZ
- Nigel Stallard, Ph.D., Senior Research Fellow, Medical and Pharmaceutical Statistics Research Unit, The University of Reading, Early Gate Reading, United Kingdom
- Arthur A.J. van Iersel, Ph.D., Senior Toxicologist, National Institute of Public Health and the Environment, Centre for Alternatives to Animal Testing, Bilthoven, The Netherlands
- Gary Wnorowski, B.S., Laboratory Director, Product Safety Labs, East Brunswick, NJ

The following ICCVAM agency representatives were present:

- Dr. George Cushmac (Acute Toxicity Working Group; ATWG), U.S. Department of Transportation
- Dr. Kailash Gupta (ATWG), Consumer Product Safety Commission
- Dr. David Hattan, Center for Food Safety and Applied Nutrition (CFSAN), Food and Drug Administration
- Dr. Richard Hill, (ICCVAM Co-Chair & ATWG), U.S. Environmental Protection Agency
- Ms. Vera Hudson, National Library of Medicine
- Dr. Devaraya Jagannath, Center for Veterinary Medicine (CVM), Food and Drug Administration
- Dr. William Stokes (ICCVAM Co-Chair & ATWG), National Institute of Environmental Health Sciences
- Dr. Kenneth Weber, National Institute for Occupational Safety and Health
- Dr. Errol Zeiger, National Institute of Environmental Health Sciences

The following members of the ICCVAM Acute Toxicity Working Group (ATWG) were present:

- Dr. Byron Backus, U.S. Environmental Protection Agency
- Mr. David Farrar, U.S. Environmental Protection Agency
- Dr. Roger Gardner, U.S. Environmental Protection Agency
- Dr. Masih Hashim, U.S. Environmental Protection Agency
- Dr. Elizabeth Margosches, U.S. Environmental Protection Agency
- Dr. Jeanie McAndrew, U.S. Environmental Protection Agency
- Dr. Debbie McCall, U.S. Environmental Protection Agency
- Dr. John Redden, U.S. Environmental Protection Agency
- Dr. Amy Rispin, U.S. Environmental Protection Agency
- Dr. Roy Sjoblad, U.S. Environmental Protection Agency
- Dr. Patrick Swann, Food and Drug Administration

The following members of the NICEATM Staff were present:

- Ms. Loretta Frye, National Institute of Environmental Health Sciences
- Mr. Brad Blackard, ILS, Inc.
- Ms. Sue Brenzel, ILS, Inc.
- Dr. Finis Cavender, ILS, Inc.
- Dr. Tom Goldsworthy, ILS, Inc.
- Ms. Christina Inhof, ILS, Inc.
- Ms. Linda Litchfield, ILS, Inc.
- Dr. Barry Margolin ILS, Inc.
- Dr. Ray Tice, ILS, Inc.

The following members of the public were present:

- Diane Beal, U.S. Environmental Protection Agency
- Dr. Gregg Carr, Procter and Gamble
- Eric Wilson, People for the Ethical Treatment of Animals (PETA)
- Jacqueline Russell, U.S. Environmental Protection Agency
- Nicholas Mastrota, U.S. Environmental Protection Agency
- Carolyn Lingemen, Bethesda Environmental Health
- Monica Vegarra, Covance
- Martin Stephen, Humane Society of the U.S.
- Dr. Katherine Stitzel, Procter and Gamble
- Merrill Tisdel, Novartis
- Ann Marie Gebhart, UL
- Jean Holmes, U.S. Environmental Protection Agency
- Debbie Vich, DuPont
- Carol Finlay, DuPont
- Penny Fenner-Crisp, U.S. Environmental Protection Agency
- Roy Sjoblad, U.S. Environmental Protection Agency
- W.T. Meyer, U.S. Environmental Protection Agency
- Susan Makris, U.S. Environmental Protection Agency
- Jeff Ferguson, Rohm & Haas
- Sara Thurin Rollin, Bureau of Natural Affairs, Inc. (BNA)
- Lee Hofmann, U.S. Environmental Protection Agency
- Mario Styliano, U.S. Environmental Protection Agency
- Andrew Rowan, Humane Society of U.S.
- Liesel Wolff, PETA

Introductions

Dr. Curtis Klaassen, co-chair, called the meeting of the Special Emphasis Panel (SEP) for the independent peer review of the revised UDP to order at 8:36 a.m. and asked each attendee to state their name and affiliation. Dr. Klaassen informed the participants that the public would be given the opportunity to speak, that each speaker from the public would be limited to seven (7)

minutes, and that anyone addressing the group to please state their name for the benefit of the transcriptionist.

Dr. William Stokes read the Statement of Conflict of Interest and explained policies and procedures regarding confidentiality and avoidance of conflict of interest situations.

Welcome from the Acting Director of the Environmental Toxicology Program, NIEHS

Dr. Chris Portier thanked the ICCVAM participating agencies and the peer review panel (Panel) for their efforts. He presented an overview of the National Toxicology Program (NTP) and delineated several NTP initiatives associated with alternatives to traditional toxicity testing, including toxicogenomics, transgenic models, structure activity relationships, and mechanism-based mathematical modeling and computer simulation.

Introduction to ICCVAM and the ICCVAM Test Method Review Process

Dr. William Stokes, ICCVAM Co-Chair and Director of NICEATM, presented the background and history of ICCVAM and NICEATM and the process and procedures for evaluation of the Up-and-Down Procedure. He discussed the role of the ICCVAM Committee, its expert subgroup (Acute Toxicity Working Group [ATWG]), the Panel, and Public Law 103-43. This law directed the NIEHS to develop and validate alternative methods that can reduce or eliminate the use of animals in acute or chronic toxicity testing, establish criteria for the validation and regulatory acceptance of alternative testing methods, and recommend a process through which scientifically validated alternative methods can be accepted for regulatory use.

Criteria and processes for validation and regulatory acceptance were developed in conjunction with 13 other Federal agencies and programs with broad input from the public. These are described in the document "Validation and Regulatory Acceptance of Toxicological Test Methods: A Report of the Ad Hoc Interagency Coordinating Committee on the Validation of Alternative Methods," NIH Publication 97-3981, NIEHS, 1997. This document is available on the internet at:
http://iccvam.niehs.nih.gov/docs/guidelines/validate.pdf

NIEHS and 13 other Federal regulatory and research agencies and programs subsequently established ICCVAM in a collaborative effort. The Committee's functions include the coordination of interagency reviews of toxicological test methods and communication with stakeholders throughout the process of test method development and validation, keeping in mind the 3 Rs (refinement, reduction, and replacement) of animal use.

The following Federal regulatory and research agencies and organizations are participating in this effort:

- Consumer Product Safety Commission
- Department of Defense
- Department of Energy

- Department of Health and Human Services
  - Agency for Toxic Substances and Disease Registry
  - Food and Drug Administration
  - National Institutes of Health
    - National Cancer Institute
    - National Institute of Environmental Health Sciences
    - National Library of Medicine
- National Institute for Occupational Safety and Health/Centers for Disease Control and Prevention
- Department of the Interior
- Department of Labor
  - Occupational Safety and Health Administration
- Department of Transportation
  - Research and Special Programs Administration
- Environmental Protection Agency

Independent peer review is an essential prerequisite for consideration of a method for regulatory acceptance (NIEHS, 1997). The Panel was charged with evaluating and developing a consensus on the usefulness and limitations for each of the tests described in the UDP (Primary Test, Limit Test, and Supplemental Test) as a replacement for the OECD TG 401. The proposed test method and results of the peer review will be forwarded by ICCVAM to federal agencies for consideration. Federal agencies will determine the regulatory acceptability of the method according to their mandates.

Summary of Current Agency Requirements for Acute Oral Toxicity Data

Dr. Amy Rispin spoke on behalf of regulatory agencies regarding the needs for acute toxicity information for hazard classification and labeling and risk mitigation in the U.S. She presented an overview of the history and current agency regulations with regard to acute toxicity testing guidelines. Dr. Rispin stated that in 1999, OECD agreed that TGs 420, 423, and 425 should be updated and refined to meet the regulatory needs of member countries. These methods should include determination of slope, confidence intervals, and data to support classification and/or assessment of acute toxicity at a minimum of 5 mg/kg and a maximum of 5,000 mg/kg. Additionally, OECD member countries have been involved in international negotiations to characterize a harmonized scheme of classification for all health effect endpoints, to encourage the use of single sexes in testing, to take advantage of sequential dosing, to utilize appropriate statistical methods in these alternative guidelines, and to incorporate and use data from well-designed sighting studies.

The revisions to the UDP were intended to improve the performance of the basic UDP for a variety of chemicals and implement the recommendations made at a March 1999 OECD meeting in which discussions were aimed at alternative methods to TG 401. With increased dosing intervals, the Primary Test in the revised UDP method functions both as a range-finding test and a main test. With the revision, the limit dose of the test was extended to 5,000 mg/kg and sequential dosing was incorporated into all three tests (Primary, Limit, and Supplemental). Dr. Rispin added that the starting dose levels were evaluated to ensure that the test performed well

with new globally harmonized classification limits. Complementary testing can be conducted for slope and confidence intervals, by using the results of the Primary Test and the Supplemental Test. Additionally, the latest humane practices for animal handling and testing were incorporated.

Overview of the Revised Up-and-Down Procedure

Dr. Katherine Stitzel described the three test procedures (Primary, Limit, and Supplemental) outlined in the UDP guideline and provided background on the revised UDP procedure. She explained that the UDP is more useful when a point estimate of LD50 or an estimate of slope is needed, and discussed the requirements for achieving a point estimate of the LD50. The Primary Test provides an estimate of the LD50, the Limit Test indicates whether the LD50 is above or below the limit dose, and the Supplemental Test estimates the slope and the confidence interval. Information on the three test procedures may be found in the UDP background review document (BRD) and other supporting materials on the internet at http://iccvam.niehs.nih.gov/methods/udpdocs/AllBRDlk.pdf

Panel Presentations on Protocol and Tests of the UDP

Dr. Curtis Klaassen stated that the meeting would proceed with presentation of reports from the four sections charged with evaluation of the UDP: General Protocol Considerations, the Primary Test, Limit Test, and Supplemental Test Sections.

General Protocol Considerations

Dr. Janice Kuhn, the section coordinator, reviewed the general protocol for the three tests (Primary, Limit, and Supplemental). Section members included Ms. Kimberly Bonnette and Mr. Gary Wnorowski.

Dr. Kuhn explained that the role of this section was to offer a practical, laboratory-based perspective to the UDP. The assigned tasks were to evaluate the protocol, the level of ambiguity in the guideline, the practicality of the guideline in a laboratory setting, and the possibility of obtaining acute toxicity information without incurring undue increases in time or expense.

The Section concluded that the proposed test method protocol was generally adequate, but recommended the following additions and/or changes:

- The use of either sex (all males or all females) should be permitted unless information is available suggesting that one sex is more sensitive;
- The use of constant volume or constant concentration of the test material during administration should be allowed;
- All reference to littermates should be excluded from the Guideline;
- Animals of 8 to 12 weeks of age should be used;
- Individual animal body weights on the day of dosing must be within 20% of the mean body weight for all animals dosed throughout the study;

- Additional guidance that incorporates how to use all pre-start data (e.g., *in vitro* test results, physical and chemical properties) should be provided in the Guideline;
- The overall usefulness of information (e.g., clinical signs, time course of effects, target organs, pathology, etc.) gained beyond the LD50 should be emphasized in the Guideline; and
- The Guideline should be reorganized to improve clarity.

The conclusion of this section was that the revised UDP protocol, with minor adjustment, could replace TG 401, but that this replacement would bring an increase in costs and complexity. There was agreement with this conclusion and recommendations by the Panel members.

Revised Up-and-Down Procedure Primary Test

Dr. Wallace Hayes, the section coordinator, presented the analysis and conclusions reached by the Primary Test method reviewers, which included Drs. Bas Blaauboer, Robert Copeland, Nigel Stallard, and Mr. John Reeve.

With regard to the revised UDP Primary Test, the Section recommended that the Guideline would be improved with the following additions/revisions:

- The scientific basis should be presented in the Guideline;
- The Guideline should include a description of how historical data should be used to decide when to use the UDP Primary Test, the UDP Limit Test, or not to conduct any test;
- Additional guidance on the starting rule and a justification of the default starting dose of 175 mg/kg should be discussed in the Guideline;
- An improved description of stopping rule #3 should be included in the Guideline;
- User-friendly, validated software for test use or access to such software should be provided;
- In the Guideline, stopping rule #1 of the UDP Primary Test and the UDP Limit Test should be harmonized;
- In the Guideline, the term “half-log” units should be used throughout rather than the approximate dose progression factor of 3.2;
- Since no formal *in vivo* validation has been reported for the revised UDP Primary Test, at a minimum, a practicability evaluation of the revised test should be conducted (an appropriate working group should consider the design of this evaluation);
- In the Guideline, the overall usefulness of information (e.g., clinical signs, time course of effects, target organs, pathology, etc.) gained beyond the LD50 in the revised UDP Primary Test should be emphasized;
- The term “slope” should be defined in the Guideline; and
- The Guideline should state that any suitable statistical LD50 estimate method (e.g., isotonic regression) may be used.

The conclusion of this Section was that the revised UDP Primary Test would provide the same and possibly additional information when compared to TG 401, and that the Primary Test can replace TG 401 for classification purposes with the use of fewer animals. There was agreement with this conclusion and recommendations among the Panel members.

Public Comment Session

Mr. Mario Stylianou from the National Institutes of Health, the National Heart, Lung and Blood Institute described an additional method of estimating the LDp by using the maximum likelihood method modified as an isotonic regression estimate. When using the modified isotonic estimate, no estimate of *sigma* is needed. He stated that the use of the modified isotonic estimate also provided an estimate of the dose-response curve and that utilization of a statistical program reduces the level of complexity.

Dr. Andrew Rowan of the Humane Society of the U.S. stated that the assumption that the LD50 is a necessity was discouraging and that no precision exists with the LD50. Dr. Rowan challenged the Panel to determine the underlying assumptions that this test method is better than the previous and that the results are accurate.

Ms. Liesel Wolf of PETA (People for the Ethical Treatment of Animals) read a written commentary on behalf of Mary Beth Sweetland, the director of research investigations and the vice president of PETA. These written comments are included as an appendix. Ms. Wolf stated that the U.S. EPA remains one of the main obstacles to the OECD deletion of the *in vivo* LD50 test.

Dr. Martin Stephens, Humane Society of the U.S., stated that animal protectionists were concerned with the number of animals needed for the Revised UDP and that the quest for precision seemed more important than the protection of animals. He expressed concern over the males being bred and not used for testing and that the maximum dose level was increased from 2,000 to 5,000 mg/kg, thereby increasing distress levels in animals. Further concern was expressed with starting at high dose levels and then subsequently decreasing the dose levels. Dr. Stephens also called on the Federal agencies to provide information to interested parties on the extent of testing conducted.

Revised Up-and-Down Procedure Limit Test

Dr. George Alexeeff, section coordinator, presented the analysis and conclusions reached by the test method performance section reviewers, which included Drs. A.A.J. van Iersel and Robert Condon.

With regard to the revised UDP Limit Test, the Panel recommended that:

- The scientific basis and rationale should be added to the Guideline;
- Additional discussion of the applicability of the revised UDP Limit Test in the strategy of hazard or safety assessment should be included in the Guideline (a flow chart with decision criteria covering the complete testing scheme might be an efficient way to attain this goal);
- Consideration should be given to reorganizing the Guideline to improve clarity;
- Clarification of the selection of the limit dose would be helpful in the Guideline and the BRD;
- Additional calculations to justify the benefits of the revised UDP Limit Test would be helpful (i.e., the document should provide probability estimates for accuracy using criteria that

compare the revised UDP Limit Test to OECD TG 401 to clearly delineate the benefits, and the document should provide probability estimates for accuracy using more stringent criteria to determine if a further reduction in the number of animals tested is possible);

- The value of the revised UDP Limit Test would be improved if additional calculations were conducted regarding the probability for correct classification using other decision criteria; and
- The different stopping rules for the upper limit dose in the revised UDP Primary and Limit Tests may cause confusion and additional explanation in the BRD is suggested to address this issue.

The conclusion of this Section was that the Limit Test may be performed when it is necessary to determine if the LD50 is above a defined limit (2,000 or 5,000 mg/kg). There was agreement with this conclusion and recommendations among the Panel members.

Supplemental Test

Dr. Bob Scala, the section coordinator, presented the analysis and conclusions reached by the supplemental test section reviewers, which included Drs. Nancy Flournoy, Phil Botham, Wyman Dorough, and Charles Hastings.

With regard to the UDP Supplemental Test, this Section recommended that:

- Regulatory data needs currently addressed by estimation of the slope and confidence interval derived from acute oral toxicity studies in the rat and other species need to be more clearly defined; and
- Consideration should be given as to whether the slope and confidence interval are the most appropriate parameters for risk assessment or whether risk assessment needs can be addressed more directly. For example, if estimates of points on the dose-response curve well below the median lethal dose are needed in environmental risk assessment, more efficient methods should be considered.

The UDP Supplemental Test for slope and confidence interval was not recommended for adoption. The Panel concluded that they were unable to evaluate the utility of the test because sufficient information regarding the use of the resulting data was not provided.

Peer Review Panel Conclusions

Co-chairperson, Dr. Diane Gerken, led the discussion and voting regarding the two major questions posed to the Panel.

The Panel was charged with separately addressing the following two questions for each of the three tests:

1. Has the revised UDP been evaluated sufficiently and is its performance satisfactory to support its adoption as a substitute for the currently accepted UDP

(OECD, 1998), and as a substitute for the traditional LD50 test for acute oral toxicity (U.S. EPA Health Effects Guidelines, OPTTS 870.1100; OECD, 1987)?

2. With respect to animal welfare, does the revised UDP adequately consider and incorporate where scientifically feasible, procedures that refine, reduce, and/or replace animal use?

In response to these questions, the Panel concluded that:

- The performance of the revised UDP Primary Test is satisfactory and exceeds the performance of OECD TG 401 in providing, with fewer animals, both an improved estimate of the LD50 for the purpose of hazard classification and more accurate information on acute toxicity. In particular, the use of 0.5 log units for dose spacing is reasonable and appropriate based on experience and the results of computer simulations. Three disadvantages of the revised UDP Primary Test recognized by the Panel were: a) the increased length of time needed to conduct a study; b) the increased costs per test material evaluated; and c) the increased complexity of the protocol.

- The revised UDP Limit Test at 2000 or 5000 mg/kg is expected to perform as well as or better than the Limit Test in OECD TG 401, with a reduction in the number of animals needed to conduct a test.

- The UDP Supplemental Test for slope and confidence interval is not recommended for adoption. The Panel was unable to evaluate the utility of the test because sufficient information regarding the use of the resulting data was not provided. As a consequence, any impact on animal use was not assessed.

The revised UDP Primary Test and the revised UDP Limit Test will reduce the number of animals used, but will not replace the use of animals. The Panel could not reach a consensus on the overall issue of refinement. However, the OECD Guidance Document on the Recognition, Assessment, and Use of Clinical Signs as Humane Endpoints for Experimental Animals used in Safety Evaluation (OECD, 1999), referenced in the revised UDP Guideline, provides an element of refinement.

Dr. Stokes on behalf of ICCVAM and its participating agencies thanked the Panel for their thoughtful deliberations and careful evaluation of the test method and background materials.

Dr. Klaassen adjourned the meeting at 5:10 p.m.

Interagency Coordinating Committee on the Validation of Alternative Methods (ICCVAM)

Independent Peer Review of the Revised Up-and-Down Procedure (UDP) for Acute Oral Toxicity

Tuesday, August 21, 2001, 10:00 a.m. – 12:00 p.m. EDT

UDP Peer Review Panel
Teleconference Call Minutes

National Institutes of Environmental Health Science (NIEHS)

79 T.W. Alexander Drive

Building 4401, Room 3162

Research Triangle Park, North Carolina

Teleconference Attendees:

UDP Peer Review Panel

- Dr. Diane Gerken (Co-Chair)
  Battelle Memorial Institute
- Dr. Curtis Klaassen (Co-Chair)
  University of Kansas Medical Center
- Dr. George Alexeeff
  California Environmental Protection Agency
- Dr. Phil Botham
  Syngenta, Ltd.
- Dr. Robert Condon
  Consulting Biostatistician
- Dr. Robert Copeland
  Howard University
- Dr. Nancy Flournoy
  American University
- Dr. A. Wallace Hayes
  The Gillette Company
- Dr. Janice Kuhn
  Stillmeadow, Inc.
- Mr. John Reeve
  New Zealand Ministry of Agriculture and Forestry
- Dr. Robert Scala
  Toxicology Consultant
- Mr. Gary Wnorowski
  Product Safety Labs

UDP Technical Task Force Representatives

- Mr. David Farrar
  U.S. Environmental Protection Agency
- Dr. Michael Green
  Consumer Product Safety Commission
- Dr. Kailash Gupta
  Consumer Product Safety Commission
- Dr. Elizabeth Margosches
  U.S. Environmental Protection Agency
- Mr. William Meyer
  U.S. Environmental Protection Agency
- Dr. Amy Rispin
  U.S. Environmental Protection Agency
- Dr. Kathy Stitzel
  Proctor and Gamble Co.

ICCVAM Agency Representatives

- Dr. Suzanne McMaster
  U.S. Environmental Protection Agency
- Dr. Marilyn Wind
  Consumer Product Safety Commission

NTP Interagency Center for the Evaluation of Alternative Toxicological Methods (NICEATM)

- Dr. William Stokes (ICCVAM Co-Chair)
  National Institute of Environmental Health Sciences
- Mr. Bradley Blackard
  Integrated Laboratory Systems, Inc.
- Ms. Ashlee Duncan
  Integrated Laboratory Systems, Inc.
- Ms. Loretta Frye
  National Institute of Environmental Health Sciences
- Dr. Raymond Tice
  Integrated Laboratory Systems, Inc.

General Public

- Dr. David Bombick
  R.J. Reynolds Tobacco Company
- Dr. Ian Pate
  Syngenta, Ltd.
- Mr. Andrew Ballard
  Reporter, Bureau of National Affairs

Call to Order and Introductions

Dr. Klaassen welcomed all participants and called the meeting to order at 10:20 a.m. He asked everyone to state his or her name for the record and requested that discussions be limited to Panel members only.

ICCVAM Test Method Review Process for the UDP

Dr. Stokes, co-chair of ICCVAM, thanked the Panel for their participation in the teleconference and provided background information and timelines pertaining to the UDP. He explained that the conclusions and recommendations of the Panel from the July 2000 Peer Review meeting were considered by the UDP Technical Task Force and incorporated into a revised UDP Test Guideline. The Task Force also developed a proposed procedure for calculating confidence intervals and a software program for use with the UDP. These are the items that the Panel has been asked to review during this teleconference meeting. Dr. Stokes then read the Conflict of Interest Statement; no conflicts were stated among the participants. He explained that the Panel will prepare a written report following the teleconference for publication in the UDP Peer Panel Final report, scheduled to be printed in November 2001. In accordance with Public Law 106-545, this report and accompanying ICCVAM recommendations will be forwarded to Federal agencies for consideration and action.

Peer Review Panel Discussion

Dr. Klaassen began the meeting by discussing the Panel's position on Evaluation Guidance Question #1 – *The revised draft UDP Test Guideline (June 20, 2001) incorporates modifications in accordance with the Panel's recommendations at the July 25, 2000 Peer Review Panel meeting.*
a) Are the changes consistent with the Panel's recommendations?
b) Do you concur with the revisions that have been made?

The Panel concluded that many of the requested changes had been appropriately considered and that they agreed with the changes made. However, several recommendations appeared to have not been adequately addressed in the revised UDP Test Guideline and these were considered during the teleconference on a case-by-case basis.

Recommendation: to increase flexibility and adaptability in animal use, the use of either sex or the more sensitive sex (if information is available indicating that one sex is more sensitive) should be permitted. The Panel unanimously re-affirmed this recommendation

Recommendation: the body weight of an animal on day 1 of dosing should be within 20% of the mean body weight of all previous animals used. The Panel recognized the confusion in wording in this recommendation (day 1 and previous animals) and, based on the revised language included in paragraph 14 of the revised draft Guideline, decide to withdraw this recommendation.

Recommendation: to include additional guidance for use of pre-start data (data available before the acute toxicity test is conducted) that may be helpful in determining the starting dose. The revised draft UDP Test Guideline addresses this recommendation in paragraph 4 as follows: *All available information on the test substance should be considered by the testing laboratory prior to conducting the study. Such information will include the identity and chemical structure of the substance; its physical chemical properties; the results of any other in vitro or in vivo toxicity tests on the substance or mixtures; toxicological data on structurally related substances or similar mixtures; and the anticipated use(s) of the substance. This information is useful to determine the relevance of the test for the protection of human health and the environment, and will help in the selection of an appropriate starting dose.*

Several Panel members expressed an opinion that this type of information was more appropriate for inclusion in a training session or guidance document, rather than in a guideline. Dr. Flournoy stated that the concept of this recommendation was to provide a better idea of the types of information or data to consider when selecting a starting dose level and to provide an alternative for the default starting dose level. The Panel unanimously recommended the following modification to the guideline *"All available information on the test substance should be considered by the testing laboratory prior to conducting the study. Such information may include the identity and chemical structure of the substance; its physical chemical properties; the results of any other in vitro or in vivo toxicity tests on the substance or mixtures; toxicological data on structurally related substances or similar mixtures; and the anticipated use(s) of the substance. This information is useful to determine the relevance of the test for the protection of human health and the environment. This information may be valuable in selecting a dose other than the default starting dose.*

Recommendation: that a practicability evaluation be conducted of the revised UDP Test Guideline. The Panel unanimously re-affirmed this recommendation.

Recommendation: that a separate section describing how the revised UDP Primary Test addresses reduction, refinement and replacement of animals compared to the previous tests be included in the Guideline. The Technical Task force formed the following response to this recommendation: *The Guideline significantly reduces the number of animals used in comparison to Guideline 401, which often required at least 20 animals in a test: 1) the stopping rule limits the number of animals in a test; 2) sequential dosing introduces further efficiencies in animal use; 3) initial dosing is now set to be below the LD50, increasing the percentage of animals in which dosing levels will be sub lethal and thereby providing some reduction in pain and distress; and 4) the use of a single sex reduces the number of animals needed and minimizes the variability in the test population. Theoretically using females only could lead to an oversupply of males. However, the use of male rats in animal research greatly exceeds that of females and, thus, the preference for females in acute toxicity testing may well result in a better overall balance of the use of both genders. Importantly, the guideline contains a requirement to follow the OECD Guidance Document on Humane Endpoints that should reduce the overall suffering of animals used in this type of toxicity test.*

Dr. Klaassen suggested the removal of gender specific references or the addition of the acceptability to use either gender (as per the preceding recommendation). The Panel decided to

recommend removing the gender reference (see the underlined sentences in the above paragraph) and unanimously recommended that the statement be added to the Guideline.

Recommendation: in paragraph 17a of the revised UDP Test Guideline, constant concentration should be used unless there is scientific or regulatory need for using constant volume. If constant volume is used in the performance of the UDP, concentrations used should also be supplied. The Panel unanimously recommended that this statement be added to the Guideline.

The Panel decided that all editorial recommendations for the revised UDP Test Guideline would be summarized by the Panel's co-chairs and added to the Panel's report for consideration.

Dr. Klaassen continued the deliberations by considering the Panel's position on Evaluation Guidance Question #2 ***- Is the proposed procedure for calculating a confidence interval for the LD50 appropriate and adequate for use with the revised draft UDP Test Guideline?***

Dr. Klaassen explained that the biostatisticians on the UDP Panel (Drs. Condon, Flournoy, and Stallard) had been charged with developing the Panel's position for this question. Dr. Flournoy stated that the proposed approach was interesting but, because of limitations and uncertainties with the method, the Panel statisticians felt that language should be added to the UDP Test Guideline that specifically indicates the shortcomings and limitations of the procedure. She continued by stating that as more is learned about the use of these types of statistical methods, the procedure should be modified accordingly.

Many Panel members felt that the wording in the procedure was too technical for non-statisticians to understand and the procedure was asking too much from data from so few animals. Drs. Hayes and Botham suggested that the procedure be rewritten using non-statistical language and outlining specific situations where the procedure does not perform well. Dr. Scala stated that the UDP Technical Task Force had failed to justify the need for confidence intervals and that the analysis was based on too few animals. He presented a motion to not recommend the procedure on these grounds. Dr. Hayes seconded the motion. Dr. Flournoy stated that the proposed procedure moves the field of statistics forward and, if the limitations are clearly described, should be approved by the Panel. She went on by explaining that such a procedure would always work poorly with shallow slopes. The Panel determined that situations where the procedure works poorly were not that common and as long as the limitations are described in detail, it would be appropriate to recommend.

Dr. Scala stated that he would withdraw his previous motion if the UDP Technical Task Force would rewrite the procedure to include details of its limitations. Dr. Condon added that people using the software program would not be cognizant of the limitations of the procedure and might conclude, incorrectly, that the data obtained were inadequate in situations where an infinite confidence limit was calculated by the program. He suggested that specific language be added to the software program also explaining the limitations of the confidence interval procedure.

Dr. Botham reiterated the need for an explanation of the procedure's limitations written in language that study directors would understand. The representatives of the UDP Technical Task Force agreed to work with the Panel's biostatisticians to develop these explanations.

The Panel unanimously accepted the proposed procedure for calculating confidence intervals for the LD50 as appropriate and adequate for use with the revised draft UDP Test Guideline, as long as a description of the applicability, utility, and limitations of the procedure was included in the Guideline and in the software program. The Panel biostatisticians agreed to work with the UDP Technical Task Force biostatisticians on the development of these statements, which would be circulated to the Panel for concurrence.

Dr. Klaassen continued by discussing the Panel's position on Evaluation Guidance Question #3 – *Is the software program adequate and consistent with the procedures in the revised draft UDP Test Guideline?*

The Panel unanimously agreed that the software program to accompany the UDP is adequate and consistent with the procedures in the revised draft UDP Test Guideline. Dr. Condon stated that the program may need some minor revision as related to the Panel's concerns expressed in the Question #2 discussion.

Public Comment

No public comments were made.

Peer Review Panel Conclusions and Recommendations

Dr. Klaassen briefly reviewed the conclusions and recommendations of the Panel that were voted on during the meeting.

Adjourn

Dr. Stokes again thanked the Panel members for participating in the teleconference. Dr. Klaassen adjourned the meeting at 12:30 p.m.

From: Pate Ian GBAP
Date: 10 August 2001
Attachment 1: **Direct use of the likelihood function for ED50 estimation**
Attachment 2: **Figures**
Subject: **Revised UDP Panel Comments on Evaluation Guidance Questions**

I still have one problem with the revised guidelines and this is regarding the situation with one intermediate dose.

In the case of one intermediate dose the guidelines state that the intermediate dose is to be used as the MLD. Note example 5 page 28 of the guidance document shows an example from this situation where the calculated MLD is not at the intermediate dose - how was this calculated?

Two examples below can be generated by the test process

175	0/2	0/4
550	1/4	3/4
2000	3/3	1/1
MLD	550	550
CI	381-1710	235-852

Both these data would give a point estimate of the MLD as 550. This is difficult to accept given one has a 25% response and one has a 75% response at 550. These data sets should surely be expected to give different estimates of the MLD.

The reason for this problem is covered in a **poster** I presented to the British Toxicology Society in 1989, which I have attached.

The profile likelihood function for this situation (Figure A.2, page 47 of confidence interval description document) is correctly shown as being well behaved. However, the likelihood function itself is not well behaved. My **poster figure 6** shows a three-dimensional plot of the likelihood function which has a ridge at the intermediate dose stretching to a slope of infinity. In simple terms, this is because a perfect fit can be made to the data using a step function i.e. 0% response below intermediate dose, rising from 0-100% response at the intermediate dose and then showing 100% response above it. Consequently, the maximum likelihood estimate of the MLD based on a profile likelihood will always be at the intermediate dose. The chance of a compound exhibiting this steep a dose response is minuscule in practice. For more realistic slope estimates the maximum profile likelihood will not occur at the intermediate dose (unless the observed response is 50%) but will correctly depend on the response observed.

The guidance can easily be changed to calculate the MLD by limiting the slope to the maximum practical value or by taking the mid-point of the profile likelihood confidence interval.

Direct use of the likelihood function for ED50 estimation.

I. Pate.

ICI PLC, Central Toxicology Laboratory, Alderley Park, Macclesfield, Cheshire, SK10 4TJ.

1. Introduction

The probit, and more recently the logit, model have been used for many years to relate the probability of a response to a chemical to the dose level administered. Interest usually focuses on the estimation of the dose level that is expected to produce a 50% response, the ED50. For the purposes of this poster the logit model will be examined - in practice when interest is centred around the ED50 there is little difference between logit and probit models, although logit models are simpler to handle mathematically.

The logit model can be written

$$\log \frac{p}{1-p} = -\beta (ED50 + d_i)$$

where p_i is the proportion responding at dose d_i. The model has two parameters, the ED50 and the slope β. In general, both the dose d_i and the ED50 are expressed on a $\log_{10}$ scale.

Table 1 contains a data set typical of those generated from an acute toxicity test, where the response is the death of an animal.

Table 1

Dose	Log_{10}dose	Proportion of deaths
50	1.7	0/5
500	2.7	1/5
1000	3.0	3/5
2000	3.3	5/5

The data points are plotted in Fig.1 along with a fitted logit curve (β=8,ED50=2.9). The effect on the fitted curve of varying the parameters individually is shown in Fig.2 and Fig.3. The ED50 is reduced in Fig.2a and increased in Fig.2b whilst keeping the slope constant. The slope is reduced in Fig.3a and increased in Fig.3b whilst keeping the ED50 constant. The ED50 parameter determines the position of the dose response curve relative to the dose axis whilst the slope parameter determines the steepness of the dose response curve.

2. The Likelihood Function.

Figs.1-3 demonstrate that some pairs of parameter values provide a dose response curve which fits the collected data more closely than others. The likelihood function provides a numerical measure of support for pairs of parameter values given the collected data. The likelihood function, L(ED50,), is a function of the two unknown parameters ED50 and β, and can be expressed

$$L(ED50,\beta|data) = \prod_{i=1}^{k} \{ (p_i{**}r_i)*(1-p_i){**}(n_i-r_i) \} \qquad i=1,...,k$$

where k=no. of dose levels tested and

$$P_i = \frac{\exp(-\beta(ED50+d_i)}{1 + \exp(-\beta(ED50+d_i))}$$

The values of the likelihood for the models in Fig.1-3 are shown on the legend to each plot. The best fitting model is that of Fig.1. The maximum likelihood estimates of β and ED50 are the values which maximise the likelihood function i.e. the best supported values given the data. Fig.1 shows the maximum likelihood solution for the data of table 1.

3. The Shape of the Likelihood Function.

The shape of the likelihood function is shown graphically in Figs.4-8 over a range of data sets chosen to examine

(i) the effect of sample size
(ii) the effect of of having one or no intermediate doses (i.e. doses which are not 0 or 100% responses)
(iii) the effect of 0 and 100% responses on the likelihood function.

As our interest generally lies in the ED50 estimate Figs.4b-8b show the likelihood functions rotated so as to view them from along the ED50 axis.

4. Discussion.

The purpose of this poster is to demonstrate graphically the shape of the likelihood function for the type of data common in acute toxicity tests. These are characterised by small numbers of dose levels, usually 3 or 4, with a few, usually 5, animals tested at each dose level. The data often have only one or no intermediate responses. The likelihood function can be used directly to provide point and interval estimates for the ED50.

Point estimates for the ED50 are given by the value with the greatest likelihood and are simply the classical maximum likelihood estimates. The failure of some maximum likelihood programs for data with less than two intermediate responses can be seen from Figs.6 and 7 to be caused by the indeterminancy of the slope (in both cases maximum likelihood estmates occur at a slope of infinity). Point estimates can be obtained by restricting the slope to be less than some predetermined value. This is not unreasonable biologically as an infinite slope corresponds to a dose response model in the unlikely form of a step function.

It is evident that whilst the slope is often poorly defined the range of plausible ED50 values is often tightly bounded. In addition, for data based on as few as five observations at each dose level the shape of the likelihood function is far from normal. Interval estimates which are based directly on the likelihood function, and hence take into account its shape, can be calculated for all the examples in this poster. The intervals are known as profile likelihood or likelihood ratio intervals and whilst the technical details of their calculation are beyond the scope of this poster the motivation for them is evident in Figs.4b-7b. By looking along the ED50 axis we can determine which values of the ED50 are well-supported for any value of the slope, the construction of the interval then requiring only a definition of how well-supported a value needs to be before it is placed in the "likely" interval. Details of the necessary calculations can be found in Williams (1) and Aitkin et al (2).

5. Conclusion.

Direct examination of the likelihood function can provide both point and interval estimates for the ED50 for data based on small numbers of observations at each dose levels and for data containing less than two intermediate responses. The necessary calculations can be programmed easily in GLIM (3) or FORTRAN and with a little more effort using PROC CATMOD in SAS (4).

6. References.

(1) Williams DA. Interval Estimation of The Median Lethal Dose. Biometrics, 1986; 42: 641-645.

(2) Aitkin M, Anderson D, Francis B and Hinde J. Statistical Modelling In GLIM, pp192-194. Oxford: Clarendon Press, 1989.

(3) Baker R, and Nelder,J. The GLIM System, Release 3. Oxford: distributed by the Numerical Algorithm Group, 1978.

(4) SAS Institute Inc. SAS User's Guide : Statistics, Version 5 Edition. Cary,NC: SAS Institute Inc, 1985.

Fig.4.

Dose	Log_{10}dose	Proportion of deaths
50	1.7	0/50
500	2.7	10/50
1000	3.0	30/50
2000	3.3	50/50

ED50 = 2.91 Slope = 8.0

95% Confidence Limit for ED50 = (2.85,2.96)

Fig.4 shows the likelihood function for a hypothetical data set with a large number of observations for each dose. The function is well-defined for both the ED50 and the slope, i.e. for both parameters only a narrow range of plausible values exist.

Fig.5.

Dose	Log_{10}dose	Proportion of deaths
50	1.7	0/5
500	2.7	1/5
1000	3.0	3/5
2000	3.3	5/5

ED50 = 2.91 Slope = 8.0

95% Confidence Limit for ED50 = (2.65,3.10)

Fig.5 shows the likelihood function for a data set with the same response proportions as Fig.4 but now based on only 5 observations per dose level. The likelihood function is now less well-defined, particularly in terms of the slope. The ED50 is still tightly bound although the reduction in sample size has increased the width of the confidence interval.

Fig.6.

Dose	Log_{10}dose	Proportion of deaths
50	1.7	0/5
500	2.7	1/5
2000	3.3	5/5

ED50 = 2.70 Slope = ∞

95% Confidence Limit for ED50 = (2.61,3.13)

Fig.6 shows the likelihood function for a data set with only one intermediate response. The indeterminacy of the slope seen to a small extent in Fig.5 is now grossly exaggerated, the likelihood function having a ridge at the intermediate dose. The maximum likelihood estimate occurs on this ridge at a slope of infinity. However, when we restrict interest to the ED50 the indeterminacy of the slope is not apparent and the confidence is little changedfrom that of Fig.5.

Fig.7.

Dose Log dose Proportion of deaths

Dose	Log_{10}dose	Proportion of deaths
50	1.7	0/5
2000	3.3	5/5

ED50 = (1.7,3.3) Slope = ∞

95% Confidence Limit for ED50 = (1.7,3.3)

Fig.6 shows the likelihood function for a data set with no intermediate response. The ridge in the likelihood present in Fig.6 is now apparent for all doses within the range of the 0 and 100% response. There is no unique maximum for the ED50, all values within this range are equally well-supported (at a slope of infinity).

Fig.8.

Dose	Log_{10}dose	Proportion of deaths
50	1.7	0/5
500	2.7	1/5
1000	3.0	3/5

ED50 = 2.93 Slope = 6.03

95% Confidence Limit for ED50 = (2.62,4.49)

This example is chosen to highlight the importance of the 100% response (a similar effect can be shown for the 0% response). Whilst values with 0% or 100% responses have infinite logits (or probits) they have a major impact on the likelihood function. In particular, whereas the intermediate responses define the slope of the dose response the 0% and 100% responses are critical in the definition of the location ie the ED50 estimate. The upper confidence limit for the ED50 is now considerably higher than in Fig.5.